PLOS | ONE

RESEARCH ARTICLE

Back to Water: Signature of Adaptive Evolution in Cetacean Mitochondrial tRNAs

Stefano Montelli☯, Antonella Peruffo☯, Tomaso Patarnello, Bruno Cozzi, Enrico Negrisolo*

Department of Comparative Biomedicine and Food Science, University of Padova, Legnaro (PD), Italy

☯ These authors contributed equally to this work.
* enrico.negrisolo@unipd.it

Abstract

The mitochondrion is the power plant of the eukaryotic cell, and tRNAs are the fundamental components of its translational machinery. In the present paper, the evolution of mitochondrial tRNAs was investigated in the Cetacea, a clade of Cetartiodactyla that retuned to water and thus had to adapt its metabolism to a different medium than that of its mainland ancestors. Our analysis focussed on identifying the factors that influenced the evolution of Cetacea tRNA double-helix elements, which play a pivotal role in the formation of the secondary and tertiary structures of each tRNA and consequently manipulate the whole translation machinery of the mitochondrion. Our analyses showed that the substitution pathways in the stems of different tRNAs were influenced by various factors, determining a molecular evolution that was unique to each of the 22 tRNAs. Our data suggested that the composition, AT-skew, and GC-skew of the tRNA stems were the main factors influencing the substitution process. In particular, the range of variation and the fluctuation of these parameters affected the fate of single tRNAs. Strong heterogeneity was observed among the different species of Cetacea. Finally, it appears that the evolution of mitochondrial tRNAs was also shaped by the environments in which the Cetacean taxa differentiated. This latter effect was particularly evident in toothed whales that either live in freshwater or are deep divers.

Citation: Montelli S, Peruffo A, Patarnello T, Cozzi B, Negrisolo E (2016) Back to Water: Signature of Adaptive Evolution in Cetacean Mitochondrial tRNAs. PLoS ONE 11(6): e0158129. doi:10.1371/journal.pone.0158129

Editor: Genlou Sun, Saint Mary's University, CANADA

Received: November 12, 2015

Accepted: June 11, 2016

Published: June 23, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The work was supported by a grant to AP (University of Padova ex-60% 2012) and by a grant to EN (University of Padova ex-60% 2012).

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Mitochondrial tRNAs are fundamental components of the translational machinery of the mitochondrion, which is the powerhouse of the eukaryotic cell. In most Metazoa, the mitochondrial genome (mtDNA) contains 22 tRNAs genes (hereafter named *trnX*, where X is the single letter IUPAC code for the corresponding amino acid). For most amino acids, a single tRNA represents the whole codon family. Leucine and Serine are the only known exceptions and possess two tRNAs belonging to two distinct families (i.e., *trnL1* and *tnL2*, and *trnS1* and *trnS2*, respectively). Occasionally, multiple copies of the same tRNA are present in animal mtDNAs. In this latter case, however, they are the product of duplication/multiplication processes and do not represent distinct codon families [1].

TRNAs play a key role in mitochondrial activity; thus, it is plausible that they experienced strong evolutionary constraints, particularly concerning their structural integrity, possibly further reinforced by the fact that there is a single tRNA for most amino acids.

In present paper, the evolution of mitochondrial tRNAs was investigated in the Cetacea, a clade of Cetartiodactyla that returned to the water and consequently adapted its metabolism to an environment different from that of its mainland ancestors [2,3]. Cetaceans breathe air, despite their multiple adaptations to life in water, and therefore represent a very interesting benchmark to test whether this array of adaptations has left any signatures on mitochondrial tRNAs. Cetacea is the most diverse group of current living aquatic mammals and includes 93 species [4]. A complete mtDNA genome is available for 49 species [5–17] (S1 Table). These taxa encompass all families and most of the currently recognised genera, thus providing a very good coverage of the clade.

Cetacean tRNAs have a genomic placement matching the canonical gene order of the Vertebrata mtDNA [18] (Fig 1). This means that fourteen mitochondrial tRNAs (i.e., *trnD*, *trnF*, *trnG*, *trnH*, *trnI*, *trnK*, *trnL1*, *trnL2*, *trnM*, *trnR*, *trnS1*, *trnT*, *trnV* and *trnW*) are encoded in the α strand, while the remaining eight (*trnA*, *trnC*, *trnE*, *trnN*, *trnP*, *trnQ*, *trnS2*, and *trnY*) are found in the complementary β strand [18].

Every tRNA exhibits a secondary structure, usually a cloverleaf shape, where double-helix stems are alternated with single-stranded nucleotides (Fig 1). Finally, each tRNA is further arranged in space to generate the functional, L-shaped tertiary structure [19].

The double-stranded elements found in tRNAs are the acceptor stem, the DHU stem, the anticodon stem and the TΨC stem (Fig 1). The acceptor stem is invariably composed by 7 pairs (hereafter 1-7ac-pair) of bases (Fig 1), and the DHU stem contains 3 to 4 pairs (hereafter 1-4dh-pair) and is absent from the *trnS1* of many animal species, including all those studied in this paper. The pairs in the anticodon stem usually number five (hereafter 1-5an-pair), with two exceptions in mammalian *trnS1* and *trnS2*. In both these genes, there is an extra pair located in position 5' with respect to the standard arrangement that does not have homologous structural counterparts in other tRNAs. This pair has been numbered here as the 0 pair. Finally, the TΨC stem exhibits a variable number of pairs ranging from four to six (hereafter 1-6tp-pair) (Fig 1). A general scheme is available for numbering the placement of every nucleotide in the secondary structure of a tRNA [20] (Fig 1). However, in the present work, we utilised the very simplified system presented in Fig 1, which strictly focused on the pairs found in the different stems. This alternative scheme is, in our view, much easier to follow for the general reader, who not necessarily accustomed to the more sophisticated system devised by Sprinzl et al. [20].

In present paper, the canonical Watson-Crick base pairings will be indicated by the standard dash symbol (–) (e.g., A–T). The base pairings involving G and T will be presented with a dot (•) symbol (e.g., G•T). Finally, the base pairings implying a mismatch (see below) will be described by a vertical dash (|) symbol (e.g., A|A).

Different arrangements can be observed when an identical pair is compared in the orthologous tRNAs of two species A and B. The pair is made by the same couple of bases in both taxa (e.g., A–T, G–C, and the opposite), or the pair is differently arranged in the two species. In this latter case, there are three possibilities (1–3). (1) With respect to species A, the B taxon exhibits a simple/double mismatch in the pair (e.g., A|A vs. A–T; C|C vs. A–T). In this case, the substitution/s pattern leads to a disruption of the secondary structure of the stem for that pair. Mismatches that do not prevent the formation of the cloverleaf structure or the tertiary structure are not rare in tRNAs. Mismatches can be corrected through editing processes or can persist in the tRNA stems as unusual pairings [21]. (2) With respect to species A, the B taxon exhibits the substitution of a single base of the pair (this can arise at the 5' end as well as the 3' end of the

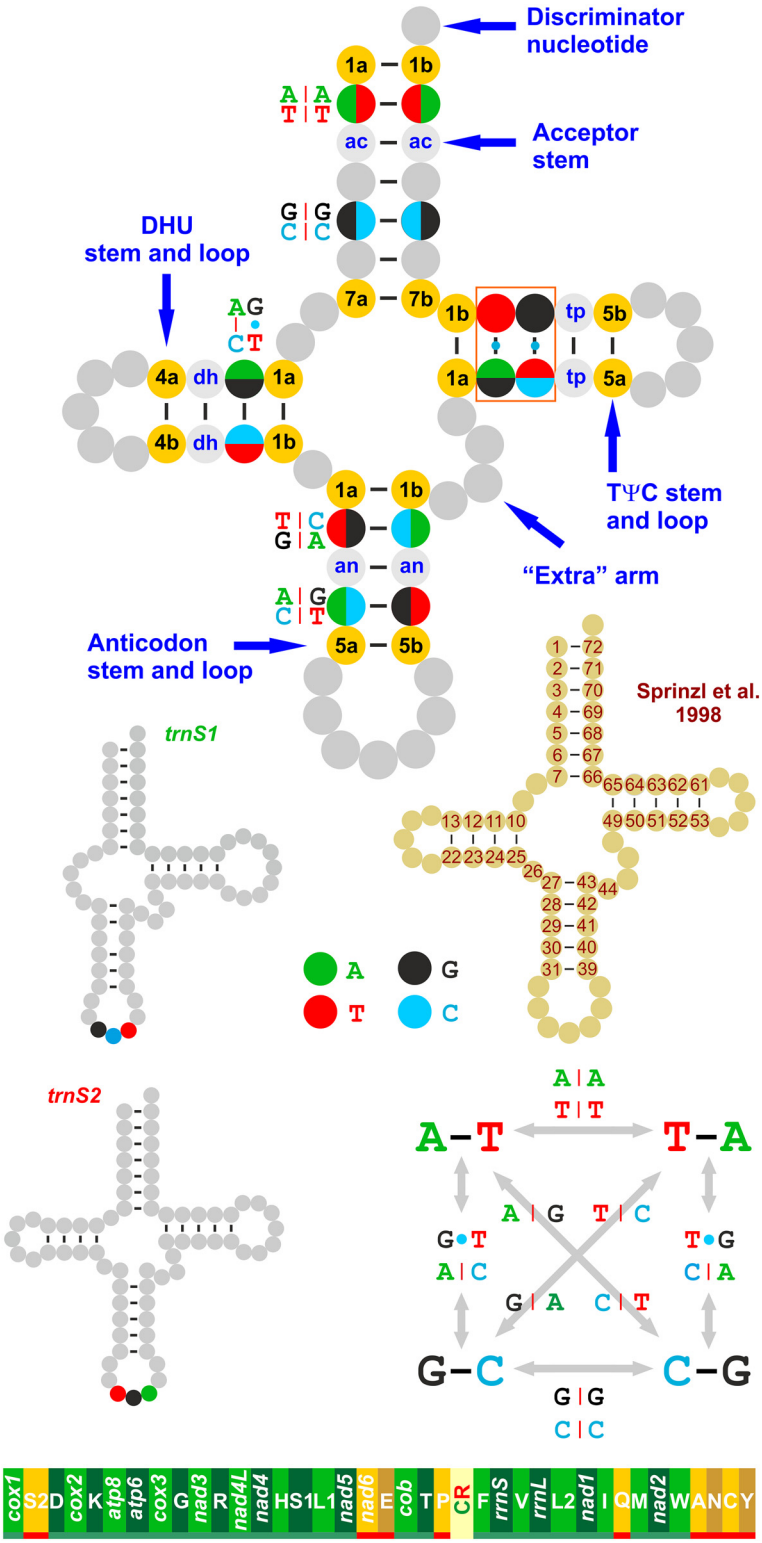


Fig 1. The tRNA nomenclature and the gene order of vertebrate mtDNA. Ac, pair of the acceptor stem; dh, pair of the DHU stem; an, pair of the anticodon stem; tp, pair of the TΨC stem. The numbering of pairs follows a 5' → 3' orientation. The 5' nucleotide of a pair is marked with a, while the 3' base is marked with b. The numbering scheme of Sprinzl et al. [20] is provided only for the stems. Base pairings are indicated as follows: –, canonical Watson-Crick base pairing; •, base pairing involving G and T; |, base pairing implying a

mismatch. A The gene order of vertebrate mtDNA is depicted at the bottom, linearised starting from *cox1*. Genes encoded on the α-strand (right to left orientation) are underlined in green, while those encoded on the β-strand are underlined in red (left to right orientation). Gene nomenclature: *atp6* and *atp8*: ATP synthase subunits 6 and 8; *cob*: apocytochrome b; *cox1-3*: cytochrome c oxidase subunits 1–3; *nad1-6* and *nad4L*: NADH dehydrogenase subunits 1–6 and 4L; *rrnS* and *rrnL*: small and large subunit ribosomal RNA (rRNA) genes; and X: transfer RNA (tRNA) genes, where X is the one-letter abbreviation of the corresponding amino acid. In particular, L1 identifies the CTN codon family; L2 the TTR codon family, S1 the AGY codon family, and S2 the TCN codon family. CR, Control Region.

doi:10.1371/journal.pone.0158129.g001

pair), which does not alter the secondary structure of the stem (e.g., G•T vs. A–T). This type of change is named hemi-compensatory base change (HCBC) [22] because its occurrence does not alter the secondary structure itself (Fig 1) [21,23,24]. (3) Finally, species B exhibits a couple of complementary bases different than species A (e.g., A–T vs. G–C). In this case, a fully compensatory base change (FCBC) occurs in species B with respect to A. This change is named fully compensatory because the substitution of both bases does not jeopardise the secondary structure integrity [22]. Two types of FCBCs exist: one (hereafter named type I) implies the substitution of a purine-pyrimidine pair with another purine-pyrimidine couple and vice versa, and the other (hereafter named type II) is characterised by a purine-pyrimidine vs. pyrimidine-purine substitution. The occurrence of type I is favoured over that of type II because type I can be produced through an intermediated HCBC. Conversely, type II requires passage through a mismatch step.

Our paper mainly focuses on identifying the factors that influenced the evolution of the elements of the tRNA double helix of Cetacea, which play a pivotal role in the formation of the secondary and tertiary structure of each tRNA and consequently modify the whole translation machinery of the mitochondrion. FCBCs, HFBCs and mismatches are globally indicated in this paper with the acronym CSBPSs (Change of Sequence in a Base Pair of a Stem).

Materials and Methods

Sequencing of the mtDNA of *Ziphius cavirostris* G. Cuvier, 1823

For the present study, we sequenced the complete mitochondrial genome of a specimen of *Z. cavirostris*. The striated muscle tissue (approximately 0.5 g) used as the starting material to extract the total DNA was obtained from a female specimen of *Z. cavirostris* that had been stored since 2007 at -80°C at the *Mediterranean marine mammal tissue bank* (MMMTB, www.marinemammals.eu) of the University of Padova (specimen # ID 135). MMMTB is a non-profit public organisation that preserves for scientific purposes the tissues of Cetacean specimens that beached and died naturally along the Italian Coasts. MMMTB promotes the study and conservation of Cetacea. MMMTB is officially supported by the Italian Ministry of Environment and is CITES credited. The scientific study of tissues obtained from MMMTB does not require the approval of an ethical committee.

The extraction was performed through a salting-out protocol [25]. The amplification and sequencing of mitochondrial DNA were performed using a mixture of mammalian universal primers [26] and primers specifically designed against available sequences belonging to the family Ziphiidae. The quality of DNA was assessed through electrophoresis in a 1% agarose gel. The PCR products were directly sequenced using the primers used for amplification. The sequencing was performed by BMR Genomics (http://www.bmr-genomics.it/; Padua, Italy). Both strands of PCR products were sequenced to ensure the standard accuracy required by this type of sequencing activity. The mtDNA consensus sequence was assembled using the SeqMan II program from the Lasergene software package (DNAStar, Madison, WI). The coverage of

the whole consensus sequence was at minimum 2X and in most cases 3X to 4X. The genome was annotated following the strategy briefly described below [27,28].

Initially, the mtDNA sequence was translated into putative proteins using the Transeq program available on the EBI website. The true identity of these polypeptides was established using the BLAST program [29,30]). The boundaries of genes were determined as follows. The 5' ends of protein-coding genes (PCGs) were defined as the first legitimate in-frame start codon (ATN, GTG, TTG, GTT) in the open reading frame (ORF) that was not located within an upstream gene encoded on the same strand. The only exceptions were *atp6* and *nad4*, which were previously demonstrated to overlap with their upstream gene i.e., *atp8* and *nad4L*, respectively, in many mtDNAs [31]. The PCG terminus was defined as the first in-frame stop codon that was encountered. When the stop codon was located within the sequence of a downstream gene encoded on the same strand, a truncated stop codon (T or TA) adjacent to the beginning of the downstream gene was designated as the termination codon. This codon was thought to be completed by polyadenylation, thereby producing a complete TAA stop codon after transcript processing. Finally, pairwise comparisons with orthologous proteins were performed using the ClustalW program [32] to better define the limits of the PCGs.

Regardless of the real initiation codon, formyl-Met was assumed to be the starting amino acid for all proteins as has been previously demonstrated in other mitochondrial genomes [33,34]).

Transfer RNA genes were identified using the tRNAscan-SE program [35] or recognised manually as sequences having the appropriate anticodon and capable of folding into the typical cloverleaf secondary structure of tRNAs [31]. The validity of these predictions was further enhanced by comparison based on multiple alignment and structural information to published orthologous counterparts [36].

The boundaries of the ribosomal *rrnS* and *rrnL* genes were those defined by the pairs of tRNAs adjacent upstream/downstream to these genes (i.e., *trnF* and *trnV* for *rrnS*; *trnV* and *trnL2* for *rrnL*).

Dataset construction

At least one complete mtDNA sequence for 49 cetacean species is currently available in GenBank (2015.09.30 release) (S1 Table). For some taxa, multiple sequences are available (e.g., *Physeter macrocephalus*). To ensure a balanced treatment of the different species, only one mtDNA sequence was included in the main dataset (see below). The only exception was *Orcinus orca*. For the killer whale, seven sequences were included, each representing one of the main clades recently identified within this taxon [13,37,38] that are possibly/probably distinct cryptic species; see de Bruyn et al. [39] for a different view. Before creating the datasets, both ingroup and outgroup mtDNA sequences were manually re-annotated to have high-quality annotated genomes. This was fundamental for identifying the correct boundaries of tRNAs. The analyses performed on tRNA evolution were very time-consuming; thus, the sequences of *Mesoplodon grayi*, *Mesoplodon ginkgodens* and *Neophocaena asiaeorientalis* became available too late to be fully implemented in our study. However, the sequences of *M. grayi and N. asiaeorientalis* were considered in some analyses (see the Results section). The complete reference dataset contained 94 taxa (listed as 94T-set in the paper). 94T-set included 46 cetaceans plus a broad selection of the main Artiodactyla lineages and two Perissodactyla. A list of taxa that were analysed in this paper is provided in S1 Table. The taxonomy of Cetacea used in the present paper follows that of Perrin [4]. The status of *Tursiops australis* as a distinct species is under scrutiny [4], but it was retained here provided that multiple mtDNA sequences exist for this taxon and were worthy of consideration, irrespective of the taxonomic validity of this species.

Multiple alignments of orthologous genes

Initially, each set of the 13 orthologous protein-coding genes derived from 94T-set was aligned using the pipeline implemented in the TranslatorX server [40]. This webtool ensures that the alignment of DNA sequences is obtained using as a backbone the multiple alignment derived from the amino acid counterparts. The MAFFT program was used to produce the alignments [41,42]). Successively, the Gblocks program (with the most stringent parameters activated) was used to the select the most conserved positions of the alignments [43]. Finally, the 13 Gblocks-processed nucleotide alignments were concatenated into a single multiple alignment (94T.13PCG.set).

The sequences of the orthologous tRNAs obtained from 94T-set were manually aligned considering the secondary structures predicted with tRNA-scan or that were available in the literature (see S1 tRNAs multiple alignments) [36]. The same strategy was applied to produce multiple alignments necessary to investigate the intraspecific variation of every tRNA for the species of cetaceans for which several/many mtDNA sequences exist. In the case of the 94T-trnXs alignment, it was not possible to model the substitution process for the most variable portions located in the DHU and TΨC loops of some tRNAs. In contrast, it was always possible to model the substitution process within the Cetacea clade.

Irrespective of the strategy used to obtain the multiple alignments, these alignments were successively imported into MEGA 5.2.2 [44] for further bioinformatic analyses.

Statistics of DNA/amino acid sequences

The AT-skew = (A-T)/(A+T) and the GC-skew = (G-C)/(G+C) were computed for the α strand of the full-length mtDNA of all 94 analysed taxa to evaluate the compositional biases [45]. The base compositions were determined with the EditSeq program from the Lasergene software package (DNAStar, Madison, WI).

The evaluation of the level of saturation in the DNA/amino acid substitution process was assessed for 94T.13PCG.set as well as for the orthologous tRNAs. In the case of 94T.13PCG.set, pairwise distances were calculated separately for whole codons, first + second positions, first positions, second positions, third positions and amino acids.

Initially, the p-distance and the maximum composite likelihood distance were calculated for each pairwise-comparison. Then, the (maximum composite likelihood distance—p-distance) difference was calculated for every pairwise comparison as a measure of the underestimation of the number of substitutions that is obtained through the p-distance. Indeed, the p-distance does not correct for possible multiple substitution events at a single position of the alignment. With no or minimal saturation, the (maximum composite likelihood distance—p-distance) difference is null or very small. In contrast, it exceeded the unit when the saturation process progressed sensibly. The (maximum composite likelihood distance—p-distance) difference was used instead of the more traditional (p-distance / maximum composite likelihood distance) ratio [46] because it allows for the calculation to be performed automatically on thousands of pairwise comparisons in a spreadsheet without the necessity of eliminating null maximum composite likelihood distances. Finally, the average (maximum composite likelihood distance—p-distance) difference was used as a global descriptor of saturation of the substitution process for every set of orthologous tRNAs.

The distances were computed with MEGA 5.2.2 [44]. The computations of the skews as well as other statistical calculations were performed using Microsoft Excel (Microsoft ™).

The total number of codons present in the whole set of Cetacea PCGs was calculated with the MEGA program. Stop codons were excluded from the calculation because they are not linked to a tRNA family. Analogously, start codons were not considered because different

codons determine the same formyl-Met as the starting amino acid [33,34]. Finally, the total number of codons belonging to each codon family was calculated, and the abundance of each codon family was expressed as the number of codons per thousand codons.

Phylogenetic analyses and the reference tree

Maximum likelihood phylogenetic analyses [47] were performed using the program RAxML 7.4.2 [48] implemented in the graphical user interface raxmlGUI 1.3.1 [49]. A nonparametric bootstrap test [50] was performed to assess the robustness of the topologies (1,000 replicates). Phylogenetic analyses were performed on nucleotide/amino acid datasets exhibiting the highest phylogenetic signals. In the case of DNA datasets, the GTR evolutionary model [51] was applied, while the heterogeneity of the substitution process was modelled with the CAT [52]. In the case of amino acid datasets, the MTMAM substitution matrix [53] was used in combination with the CAT algorithm. Partitioning schemes were used to test their effect on the tree topologies.

Phylogenetic analyses were performed on the position 2, positions 1+2, and amino-acid subsets of 94T.13PCG.set, which exhibited the highest signals, with and without partitions.

All of the obtained trees were identical to the topology depicted in S1 Fig. In the topology, most of the nodes received bootstrap support. The tree in S1 Fig was generated from the amino acid dataset. The topology revealed that many amino acids changed along the branch reaching the root of Cetacea. The mysticete *Caperea marginata* and, more markedly, the odontocetes *Kogia breviceps*, *P. macrocephalus*, *Platanista minor*, *Lipotes vexillifer*, *Pontoporia blainvillei*, *Inia geoffrensis*, and *Monodon monoceros* showed branches that were decidedly longer that those of other Cetacean species. No further details are presented here on the phylogeny of Cetacea. A comment must be introduced to explain this point. The phylogeny of Cetacea is a very active field of study, and several papers have been published on this topic [12,14,15,54–64]. The overall phylogenetic relationships among major lineages were consistently recovered in the studies mentioned above and are depicted in S1 Fig. In contrast, the vast majority of the published trees exhibit one or more points of disagreement. In the present paper, the topology of S1 Fig was used as a reference tree to map the evolution of CSBPSs. Alternative phylogenetic relationships were considered to test whether they could produce relevant changes in our results (data not shown). These topologies gave, at most, marginal variations restricted to single nodes and did not alter the global evolutionary pathway for the CSBPSs. Thus, they will not be described in detail in the present paper.

Tracking the substitution patterns of orthologous tRNAs along the reference tree

The CSBPSs occurring in the multiple alignments of orthologous tRNAs were tracked along a reference tree according to the maximum likelihood method available in MEGA 5.2.2 [44] and according to the maximum parsimony approach implemented in the Mesquite program [65]. In the latter, the nucleotide changes were assumed to be unordered events. The mismatches occurring at the boundaries between DHU and TΨC arms and loops were not considered. This choice was dictated by the fact that in some cases, the length of the arms was variable without disrupting the secondary structure (S1 tRNAs Multiple Alignments).

Results

Introductory note. Here, only the main results are provided. A more detailed description is presented in S1 Extended Results.

The mitochondrial genome of *Ziphius cavirostris*

The mtDNA of a specimen of *Z. cavirostris*, sequenced for this paper, is briefly described here. The new mitochondrial genome was 16,352 bp long. This value was very close to the average value obtained for the dataset analysed in the present work (16,436 ± 124). The *Z. cavirostris* genome contained the 37 genes almost universally found in animal mtDNAs i.e., 13 PCGs, two ribosomal rRNAs and 22 tRNAs. The gene order was typical for vertebrate mtDNAs (Fig 1), with 28 genes encoded on the α-strand and nine present on the opposite β-strand. Most of the PCGs started with ATG and ended with TAA or the incomplete stop codons TA(a) and T(aa). The genes on the same/opposite strand overlapped, were contiguous or were separated by intergenic spacers encompassing a variable number of nucleotides (S2 Fig). The mtDNA sequence of *Z. cavirostris* is available in EBI/GenBank under accession number LN997430.

Occurrence of CSBPSs in the tRNAs of Cetacea

The computation of the p-distance and (maximum composite likelihood distance—p-distance) difference allowed the level of conservation and the possible underestimation of the substitution patterns in Cetacea tRNAs (S3 and S4 Figs, respectively) to be determined. The most conserved tRNA was *trnG*, and the most variable was *trnH*. The minimum and maximum values for the (maximum composite likelihood distance—p-distance) difference were observed in *trnE* and *trnH*, respectively. The (maximum composite likelihood distance—p-distance) difference values demonstrated that the observed CSBPSs did not grossly underestimate the true number of CSBPSs in the tRNAs.

A total of 603 CSBPSs (136 FCBCs, 320 HCBCs, and 147 mismatches) were identified in the tRNAs of Cetacea (Fig 2; S1 tRNA Multiple Alignments).

Most of the FCBCs were of type I (131), and only four were of type II. T–A vs. A–T occurred on the 5an-pair of *trnH*. G–T vs. T–A was found on the 7ac-pair of *trnQ*. C–G vs. A–T was identified on the 2tp-pair of *trnE*. A–T vs. T–A was detected on the 2tp-pair of *trnN* (Fig 2; S1 tRNA Multiple Alignments).

The number of FCBCs detected in a single tRNA ranged from 1 to 18 (Figs 1 and 2). Ten or more FCBCs were found only in α-strand-encoded tRNAs (Figs 1 and 2). Most of the β-strand tRNAs exhibited four or fewer FCBCs. The bases alternating on the type I FCBCs followed three patterns (a-c). (a) Only purines alternated at the 5' end of the pair, and only pyrimidines occurred at the 3' end (e.g., *trnD*). (b) Both purine/pyrimidine bases occurred in the substitution at the 5' and 3' ends of the pair (e.g., *trnA*) with a variable prevalence of the first/second type of base. (c) Only pyrimidines alternated at the 5' end, and only purines occurred at the 3' end of the pair (e.g., *trnC*) (S1 tRNA Multiple Alignments).

The 320 HCBCs were distributed mainly into four symmetrical types (Fig 2; S1 Extended Results). The number of HCBCs in a single tRNA ranged from 0 (*trnN*) to 35 in (*trnE*). The distribution of HCBCs was β-strand biased. Indeed, the α-strand tRNAs showed 144 HCBCs, while the β-strand tRNAs exhibited 176 HCBCs (Fig 2).

The 147 mismatches belonged to 35 different types (Fig 2; S1 Extended Results) and ranged from 0 (e.g., *trnP*) to 30 (*trnR*). The mismatches exhibited an uneven and α-strand-biased distribution (Fig 2, S1 Extended Results).

Patterns of CSBPS distribution in Cetacea tRNAs

The distribution of CSBPSs was very variable in the tRNAs (Fig 2; S1 Extended Results). FCBCs, HCBCs, and mismatches were linked in their abundance (≥ 7) in several α-strand tRNAs (Fig 2). A second patter, mainly observed in β-strand tRNAs, had a high number of HCBCs coupled with a low number of FCBCs and mismatches. A low number of mismatches

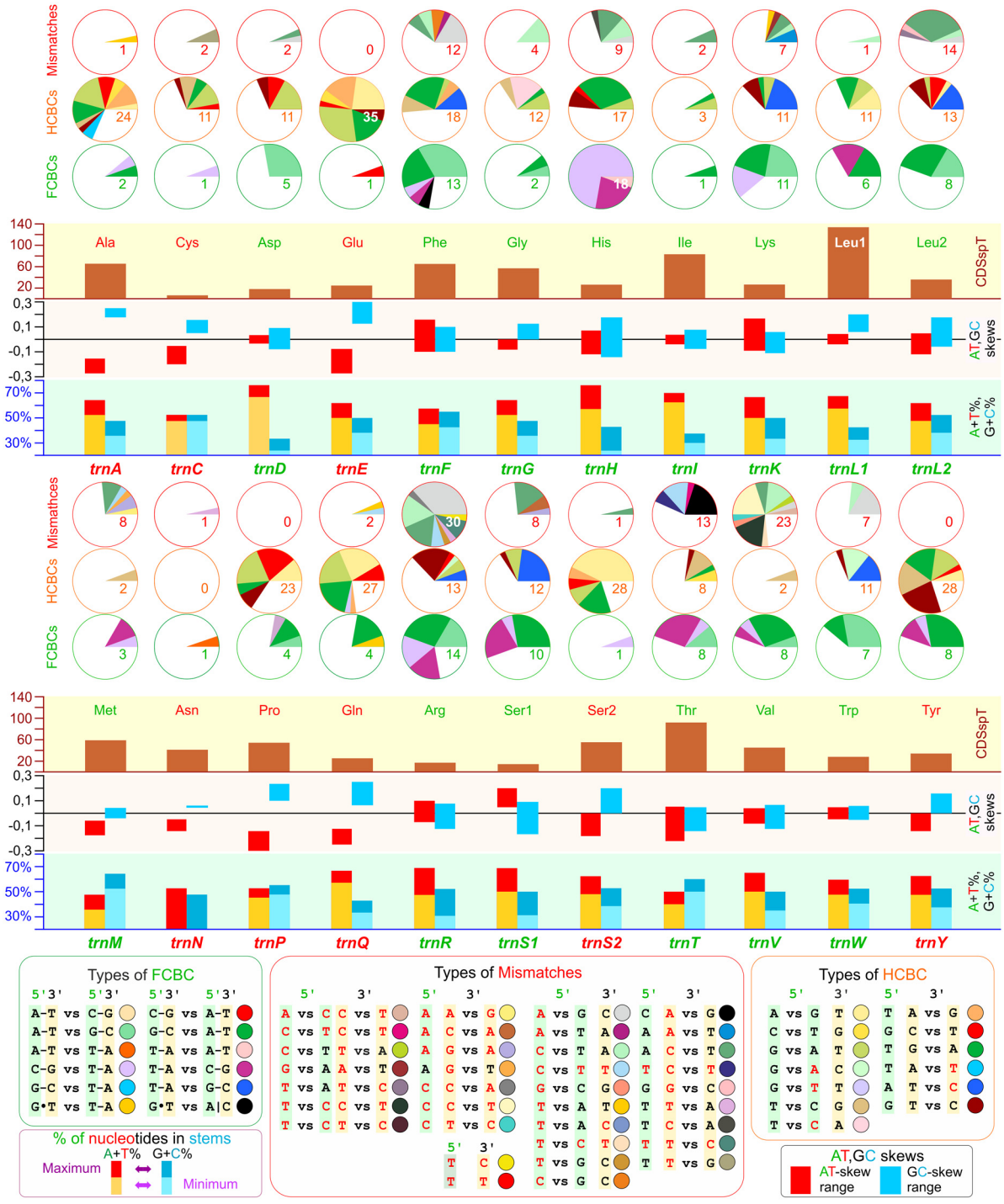


Fig 2. CSBPSs, skews, stems composition, and codon usage linked to the Cetacea tRNAs. CSBPS, change in sequence in a base pair of a stem; FCBC, fully compensatory base change; HCBC, hemi compensatory base change; Mismatch, mismatch in a base pair of a stem. The extension of the slices determining the number and type of FCBCs, HCBCs, and Mismatches for every tRNA was scaled, assuming that the whole coverage of a Pie graph was reached only in the tRNA exhibiting the maximum number of in CSBPSs. This approach allows, in our view, to better appreciate the variation of CSBPSs in the various tRNAs. CDSpT, codons per thousand codons associated to a tRNA. A+T%, percentage of A+T in the stems; G+C%, percentage of G+C in the stems. For A+T% and G+C%, the maximum and minimum values are provided. AT- and GC-skews, skews calculated for the stems of tRNAs. For AT- and GC-skews, the range of variation is provided for every tRNA.

doi:10.1371/journal.pone.0158129.g002

coupled with a moderate number of FCBCs and a higher number of HCBCs was present in *trnD*, *trnL1*, and *trnY*. A small number of base changes characterised *trnI* and *trnN*. Few FCBCs and HCBCs and a good number of mismatches occurred in *trnM*. Finally, *trnV* exhibited a high number of mismatches coupled with a good number of FCBCs and a low number of HCBCs (Fig 2; S1 tRNAs multiple alignments).

Factors influencing the occurrence and type of CSBPSs

The total number of codons encoded by the cetacean mtDNAs as well the number of codons for each codon family were homogenous (Fig 2; S1 Extended Results). The number of FCBCs, HCBCs, and mismatches were not linked to the abundance of codon families (Fig 2; S1 Extended Results).

Globally, the occurrence of CSBPSs in different tRNAs was influenced by the combined action of (a) the base content variation and (b) the asymmetrical compositional biases of the stems. These latter in several cases were opposite to the values computed for the strand encoding the analysed tRNAs. In particular, the range of variation and the fluctuation of base content, AT- and GC-skews, had a major impact on the type and abundance of CSBPSs (Fig 2; S1 Extended Results; S5 and S6 Figs).

Finally, the abundance of FCBCs, HCBCs, and mismatches did not appear to be linked to the genomic placement of different tRNAs (Figs 1 and 2). A couple of examples support this statement. *TrnA* and *trnN*, both on the β-strand and adjacent, exhibited very different behaviours. In contrast, *trnR* and *trnT*, both on the α-strand and well separated, had very similar patterns.

The stem positions associated with CSBPSs

The stem positions involved in base changes (hereafter named SPICs) were mapped and analysed in the different tRNAs (Figs 3 and 4). One, two and even all three types of substitutions were observed in the same SPIC (e.g., *trnF*, Fig 3). The number of SPICs was very variable within the 22 tRNAs (Fig 3). Due to the heterogeneity of the substitution patterns, a perfect correspondence did not exist between the percentage of SPICs and the global percentage of CSBPSs occurring in a single tRNA. Despite these vagaries, the percentage of SPICs was in good agreement with the global percentage of CSBPSs. The percentages of SPICs and FCBCs exhibited similar behaviours. Many more discrepancies existed among the percentage of SPICs and the global percentages of HCBCs and mismatches (S1 Extended Results). Each tRNA exhibited a unique pattern of SPICs and associated types of FCBCs, HCBCs, and mismatches.

The distributions of SPICs and CSBPSs are summarised in Fig 4. The occurrence of FCBCs was very variable in the pairs. Some pairs (e.g., 2ac-pair) never presented an FCBC (Fig 4a). In contrast, other pairs (e.g., 4ac-pair) were hot spots for the occurrence of FCBCs. In general, the acceptor, the anticodon, and the TΨC stems contained most of the SPICs associated with FCBCs. The DHU stem had a very limited number of SPICs associated with FCBCs (S1 Extended Results).

The SPICs associated with HCBCs were variably distributed in the different tRNAs (Fig 4a; S1 Extended Results). Some positions were heavily involved with HCBCs (e.g., the 5' end of dh1-pair), while others never exhibited an HCBC. Furthermore, the 5' end and 3' end could behave differently in the same pair (S1 Extended Results).

The SPICs associated with mismatches were more abundant on acceptor and TΨC stems. Additionally, the anticodon stem presented several SPICs linked to mismatches. Very few SPICs hosting mismatches occurred in the DHU stem. Mismatches were never detected in some positions (Fig 4a). When the global percentage of FCBCs, HCBCs, and mismatches

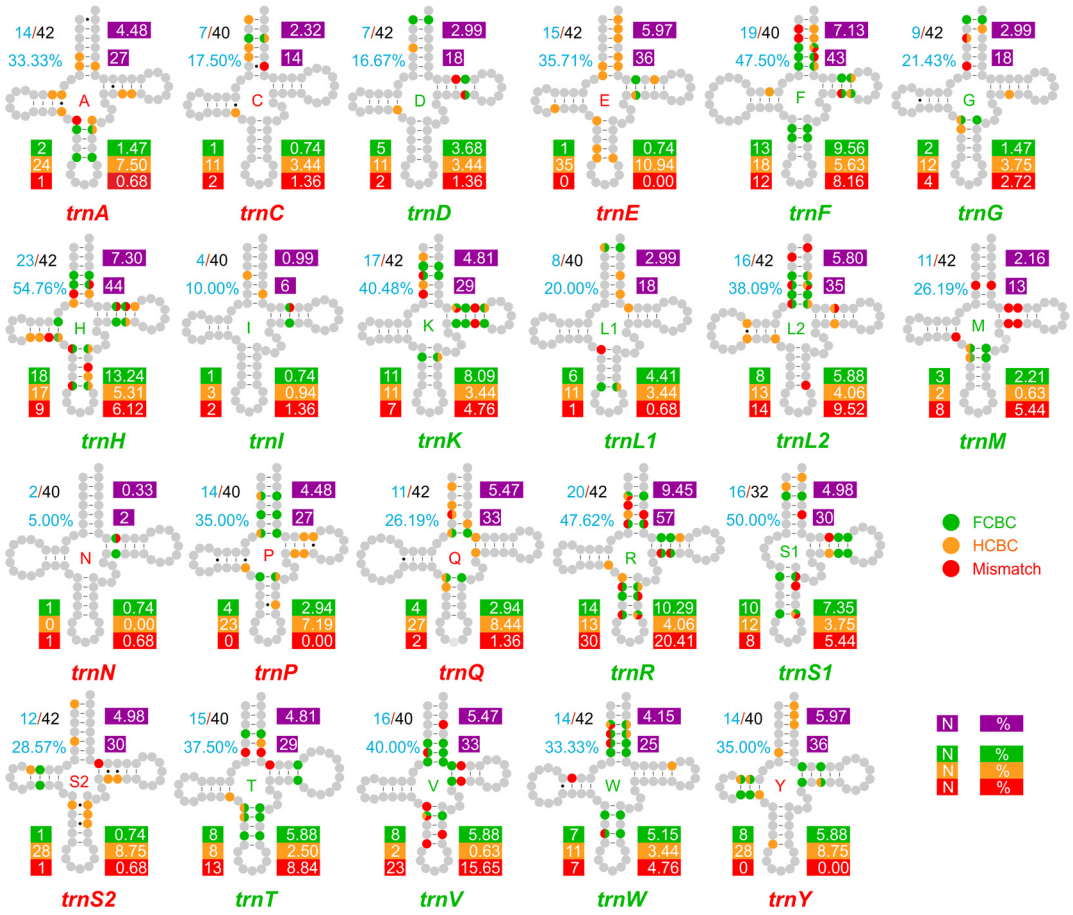


Fig 3. Distribution of CSBPSs in a single tRNA. CSBPS, change in sequence in a base pair of a stem; FCBC, fully compensatory base change; HCBC, hemi compensatory base change; Mismatch, mismatch in a base pair of a stem; SPIC, stem position involved in base change. The number/percentage of SPICs is provided in cyan. A green background is used to mark the number of FCBCs occurring in a tRNA as well as the percentage that they represent with respect to the total number of FCBCs. The orange and red backgrounds are used for the number/percentage of HCBCs and Mismatches. A purple background is used for number/percentage of CSBPSs.

doi:10.1371/journal.pone.0158129.g003

occurring at the different SPICs was evaluated, the patterns that emerged largely mirrored the abundance of SPICs described above (Fig 4b).

The occurrence of HCBCs and of mismatches exhibited an evident 5' end or 3' end distributional bias in the pairs of some tRNAs (Fig 3; S8 and S9 Figs; S1 Extended Results).

Finally, the known distribution of positions in the stems, where post-transcriptional modifications occur, was mapped and compared to the SPIC behaviour (Fig 4). A simple pattern linking these positions with CSBPSs/SPICs was not identified.

Phylogenetic distribution of CSBPSs

The distribution of CSBPSs along the reference tree is summarised in Fig 5, while the full details are provided in S7–S9 Figs.

A total of 70.32% of the CSBPSs were associated with living species of Cetacea, while the remaining 29.68% were divided among the internal nodes of the tree. In living species, the percentage of FCBCs was 66.91%, that of HCBCs was 72.19%, and that of mismatches was 69.39% (Fig 5).

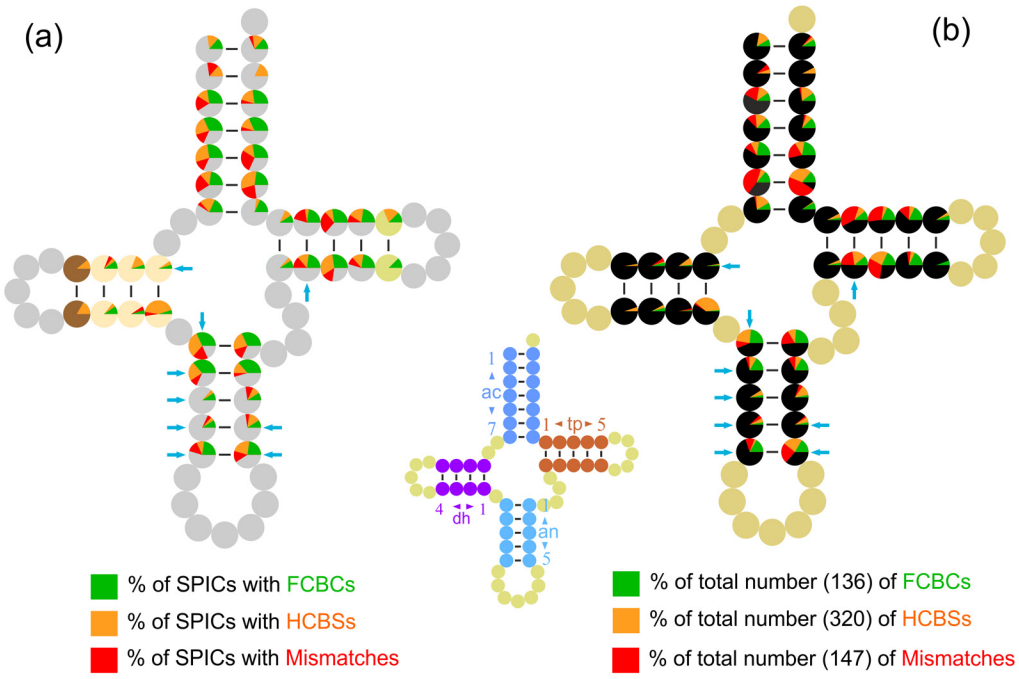


Fig 4. Global distribution of SPICs and CSBPSs in cetacean tRNAs. CSBPS, change in sequence in a base pair of a stem; FCBC, fully compensatory base change; HCBC, hemi compensatory base change; Mismatch, mismatch in a base pair of a stem; SPIC, stem position involved in base change. (a) Percent of SPICs in the whole set of tRNAs involved in CSBPSs. The extension of the slices determining the percent and type of FCBCs, HCBCs, and Mismatches was scaled assuming that the whole coverage of a Pie graph was reached only when the considered position resulted in an SPIC for the entire set of tRNAs. (b) Percent of CSBPSs occurring in a considered SPIC with respect to the total number of CSBPSs. The cyan arrows point to stem positions known to be interested in species of Mammalia by an editing activity in at least one tRNA.

doi:10.1371/journal.pone.0158129.g004

The four FCBCs of type II that were identified in Cetacea (see above) had a variable taxonomic distribution. Two were restricted to a single species i.e., *K. breviceps* (*trnN*), and *I. geoffrensis* (*trnE*) (S1 tRNA Multiple Alignments). The FCBC occurring in *trnH* appeared at the onset of Cetacea. During the cladogenetic process, successive FCBCs of type I, HCBCs and a mismatch were substituted for this CSBPS in some cetacean species (S7–S9 Figs, S1 tRNAs Multiple Alignments, *trnH*). Finally, the FCBC of type II found in *trnQ* characterised most of Odontoceti (except for *P. macrocephalus* + *K. breviceps*) and was followed by successive HCBCs.

The analysis of the distribution of the CSBPSs revealed the occurrence of a dynamic, continuous, and ongoing evolutionary mechanism of changes in the stems of tRNAs. The oldest CSBPSs were followed by successive changes (marked with an asterisk in Fig 5) that occurred in descendant groups at different taxonomic ranks (S7–S9 Figs). The CSBPSs were in several cases molecular signatures for the different clades (S1 Extended Results). In contrast, in other cases, the CSBPSs represented events of convergent/parallel evolution (S1 Extended Results). Finally, the substitution pattern produced in some cases a secondary reversion to the condition observed in outgroups due to the limited possibility of combinations of the four bases (S1 Extended Results).

The distribution of CSBPSs in living Cetacea exhibited a broad range of variation (Fig 5; S7–S9 Figs, S1 tRNA Multiple Alignments). Most of Mysticeti showed a smaller number of CSBPSs as Odontoceti. *E. robustus* and, more markedly, *C. marginata* were two exceptions to this behaviour. Within Odontoceti, most of Delphinidae exhibited a lower number of CSBPSs than

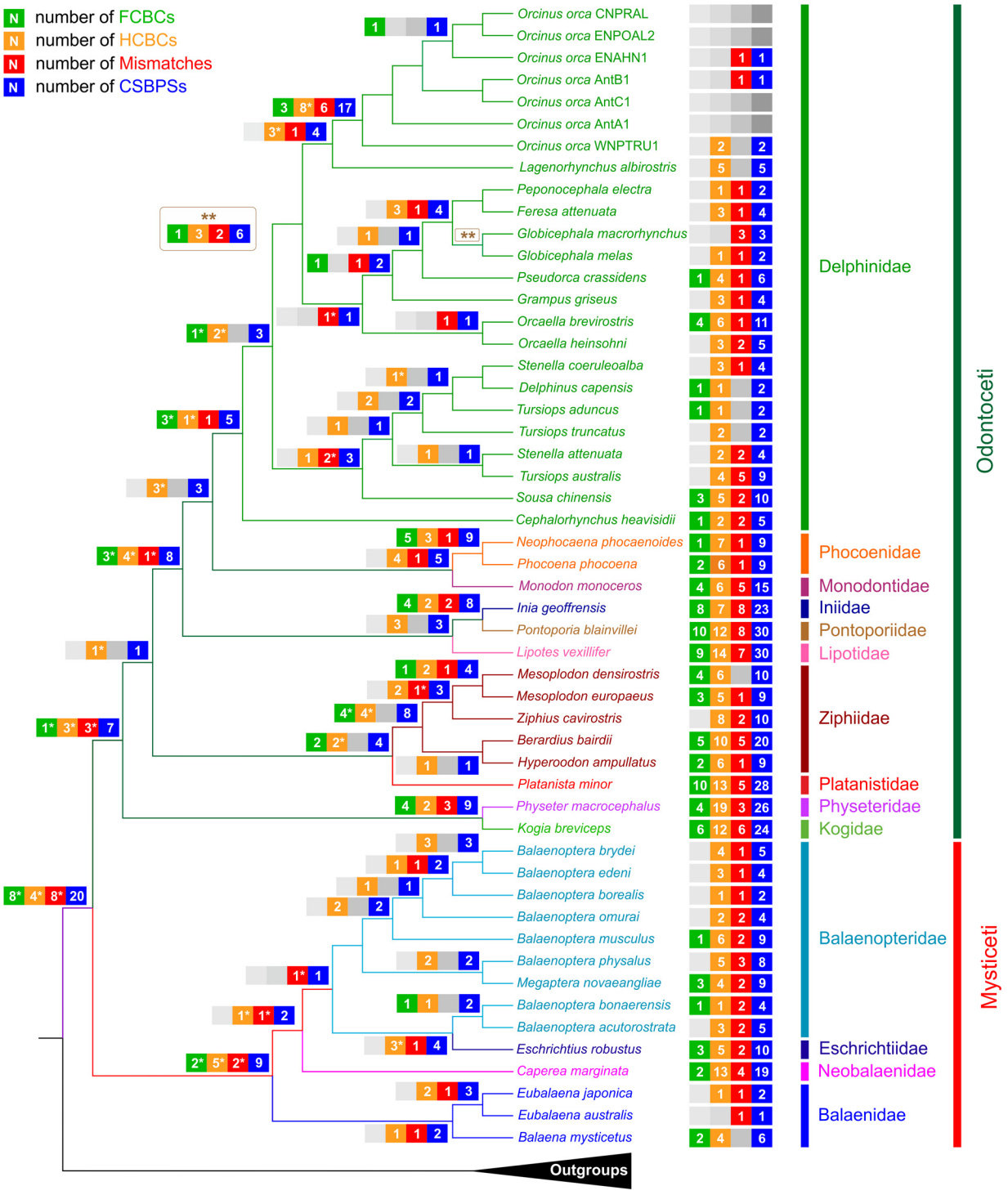


Fig 5. Mapping of CSBPSs in the Cetacea phylogenetic tree. CSBPS, change in sequence in a base pair of a stem; FCBC, fully compensatory base change; HCBC, hemi compensatory base change; Mismatch, mismatch in a base pair of a stem; Number of FCBCs, HCBCs, and Mismatches (included their sum CSBPSs) occurring at the nodes of the reference phylogenetic tree. The asterisk associated with some CSBPS values indicates that these CSBPSs were subjected to successive changes in one/some of the taxa downstream of the considered node. For details on the type of FCBCs, HCBCs, and mismatches occurring at a single node, please refer to S7–S9 Figs.

doi:10.1371/journal.pone.0158129.g005

did members of other families. *P. blainvillei* and *L. vexillifer* showed the maximum number of CSBPSs. Other species with at least 20 CSBPSs were *Berardius bairdii*, *I. geoffrensis*, *K. breviceps*, *Pl. minor*, and *P. macrocephalus*. The *O. orca* complex was a peculiar case to be analysed. If this taxon was considered an assembly of multiple cryptic species, a very low number of CSBPSs could be detected. In contrast, if different specimens of *O. orca* were considered to be derived from a single species, then a minimum of 17 CSBPSs could be assigned to this taxon (Fig 5).

A comparison of the reference topology (S1 Fig) with the distribution of CSBPSs (Fig 5) showed that a good agreement existed among the lengths of the branches and the numbers of CSBPSs.

M. grayi and *N. asiaeorientalis* were not considered in most of the analyses performed in the present paper (see Materials and Methods above). However, it was possible to include the tRNAs of these species in the multiple alignments (S1 tRNA Multiple Alignments).

The analysis of these alignments allowed us to identify for *M. grayi* at least 5 FCBCs, 3 HCBCs and 2 mismatches. Particularly interesting was the presence in the 2tp-pair of *trnN* of a type II FCBC (i.e., A–T vs. T–A). The total number of 10 CSBPSs found in *M. grayi* agreed with the values obtained for other *Mesoplodon* species (Fig 5).

Two HCBCs were unquestionably peculiar to *N. asiaeorientalis*. They were located at the 5tp-pair of *trnD* and at the 2tp-pair of *trnS2*. Furthermore, *N. asiaeorientalis* shared with *N. phocaenoides* most of the CSBPSs recorded for this latter species (Fig 5; S1 tRNA Multiple Alignments).

Intraspecific variation in CSBPSs

The intraspecific level of CSBPS variation was studied in nine cetacean species (S1 Extended Results, S2 Table). The analyses were performed for *Balaenoptera physalus*, *M. densirostris*, *Mesoplodon europaeus*, *O. orca*, *P. macrocephalus*, *Tursiops aduncus*, *Tursiops australis*, *Tursiops truncates*, and *Z. cavirostris*. Intraspecific FCBCs were not identified, regardless of the number of mtDNAs analysed (8–152). In contrast, a variable but limited number of HCBCs and mismatches was detected. The number of tRNAs containing these CSBPSs varied from 1 (*M. europaeus* and *T. australis*) to 13 (*B. physalus*). In general, the intraspecific variation in CSBPS was limited.

Discussion

Proximate causes of tRNA evolution in Cetacea

As outlined in the introduction, tRNAs are at the core of mitochondrial activity and play a key role in the synthesis of mtDNA-encoded proteins.

Our analysis focussed on the changes that occurred in the stems of 22 tRNAs because they have a major impact on the structural integrity of these molecules. It is the conservation of the structural integrity, tightly linked to the ability to properly deliver the amino acid inside the mitochondrial ribosome, that dictates the limit in variation that every tRNA can afford during its evolution.

The results in the present paper show that the 22 mitochondrial tRNAs experienced continuous change during the cladogenesis of Cetacea. However, these changes occurred in different tRNAs with very different paces in terms of both numbers and types. This finding supports the view of a complex relationship among tRNAs and several factors that can produce variation in the stem pairs. Furthermore, it is evident that a fine tuning of the actions exerted by several causes determined the diversity of effects described above.

A total of 136 FCBCs, 320 HCBCs and 147 mismatches were identified in the analysis of 94T-set. A few more CSBPSs were detected in *M*. *grayi*, *N*. *asiaeorientalis* and in the study of the intraspecific variation of selected species (see above). These latter CSBPSs did not substantially alter the main outputs derived from the analysis of 94T-set. Thus, the discussion will be focused mostly on the results obtained from the principal dataset.

Type I FCBCs dominated the evolution of cetacean tRNA stems (97%), while type II FCBCs were very rare events. The occurrence of type I FCBCs was strongly favoured and promoted by the pivotal role played by the G•T, T•G pairs involved in most HFBCs. Indeed, these pairs provide a very efficient switch to pass from A–T to G–C and vice versa without disrupting the stem integrity through two rounds of HCBCs. In contrast, a type II FCBC requires an intermediate mismatch involving a couple of identical bases, potentially hampering the stem structure or the simultaneous substitution of both nucleotides of pair with different types of bases, a very rare event.

It has been shown that A–T and G–C pairs (and their symmetrical opposites) represent the top peaks of the fitness landscape describing the evolutionary history of mitochondrial tRNAs [66]. In contrast, G•T and A|C pairs are regarded as valleys of lower fitness that must be crossed to go from one peak to the other [66]. However, the two valleys are structurally very different because G•T does not jeopardise the tRNA stem structure, while A|C hampers the stem integrity. Thus, in the fitness landscape, the G•T valley can easily be crossed. In contrast, the pathway passing through A|C to go from A–T to G–C and vice versa is much steeper. This statement is corroborated by different types of scientific evidence ranging from structural biology [23] and free energy calculations for different base pairings [24] to comparative sequence analysis and homology modelling [21] applied not only to tRNAs but to various types of RNA molecules. Finally, the valley connecting the two peaks representing the alternative pairs of a type II FCBC is very deep and difficult to be crossed, as demonstrated by the extremely limited number of these events in Cetacea. The difficulty in passing through such a valley is further corroborated by the fact that the mismatch can be almost/fully fixed, as shown in present study for *trnN*. However, a mismatch is not necessarily a defect for a stem provided that it may represent a recognition signal for molecular partners [36]. Thus, it can remain in tRNA for a long time.

Our analysis showed that the most active stem positions are involved with CSBPSs, and the positions that are affected by post-transcriptional modifications on tRNAs are linked by a complex pattern, also considering that our knowledge of the occurrence of the latter in Mammalia is rather limited [19].

The abundance of codon families does not appear to have had a major impact on CSBPS evolution. The composition of the strand encoding the different tRNAs as well as the AT- and GC-skews exerts some control on the global pattern observed for CSBPSs as previously outlined by Helm et al. [36]. However, the distribution of FCBCs, HCBCs and mismatches is much more influenced and controlled by the base composition, AT-skew, and GC-skew of the tRNA stems. Indeed, our analysis has shown that it is the range of variation of composition and skews, particularly the extent of fluctuation from positive to negative values in the tRNA stems, that deeply affects the dynamics of the observed changes. Furthermore, this behaviour can be heterogeneous even in different stems of the same tRNA and determines, at a microscale level, the occurrence and abundance of different CSBPSs. The latter can be limited to single species or extended to a variable number of taxa. A stochastic component certainly influences this process and acts at the most dynamic pairs of the stems generating the convergent/parallel evolutionary changes described above. This does not mean that tRNA stems evolve at a very fast pace, as demonstrated by the values of the (maximum composite likelihood distance

—p-distance) difference, which demonstrated that the substitution process is very far from saturation.

The type and distribution of CSBPSs is not only variable among the 22 tRNAs but also very different among different species. Thus, some CSBPSs can be proficiently used as good molecular signatures to delimit and define clades within Cetacea. The CSBPS intraspecific variation was limited, and no FCBCs were detected among the analysed species. However, the samples were small, and a much better coverage is necessary to fully assess this point.

The analysis of the distribution of CSBPSs in the Cetacea tree allowed us to hypothesise the role played by some factors in the evolutionary changes that occurred in the mitochondrial genome of Cetacea during their return to water. These evolutionary pathways are discussed in the next paragraph.

Anatomical and physiological requirements and tRNA evolution

Recent evidence [67] suggests that gene families associated with stress-responsive proteins and anaerobic metabolism are expanded in cetaceans, while genes linked to sensory receptors and body hair are contracted (for the latter, see also Nery et al. [68]). This confirms that whale and dolphin genomes reflect the physiological need for a breath-holding based metabolism and intense stress due to increased reactive oxygen species and a high-salt environment. The high energy requirements of life in the water are testified also by the parallel evolution of the *IDH2* gene, which encodes an enzyme involved in aerobic metabolism in cetaceans, primates and bats [69].

In fact, a novel expansion of a polyalanine tract of the homeobox (Hox) genes *Hoxd12* and *Hoxd13* in cetaceans implicates a selective pattern of development of the specific morphology of the thoracic limb that is transformed into the typical flipper [70]. Conversely, in contrast to that expected, the primate-dolphin comparison showed that the evolution of *microcephalin* (*MCPH1 brain-development gene*) was not associated with brain size in cetaceans [71], thus failing to pinpoint an evolutionary factor responsible for the highest brain mass in the clade of mammals.

According to our data, the distribution of CSBPSs is very variable in the different cetacean families (Fig 5). Most species of Delphinidae exhibit fewer CSBPSs than do other toothed whales. The *O. orca* complex represents the main exception to this behaviour. The smaller number of CSBPSs observed in Delphinidae has possibly been influenced by a combination of the physiological requirements of the species belonging to this family (see below) as well the relative younger age of the clade with respect to that of other Cetacea lineages [55].

Phocoenidae and Monodontidae exhibit a number of CSBPSs slightly higher than but still comparable to that of Delphinidae. The shape of the body and the general food preferences of these three families are similar and may cause the differences in the deep divers, including Kogiidae, Physeteridae and Ziphiidae, and in the estuarine and freshwater species, including Pontoporiidae, Iniidae, Lipotidae and Platanistidae. Toothed whales that live in estuarine and freshwater habitats and cetaceans that repeatedly hunt at great depths face different but equally challenging physiological stresses only partially shared by taxa living in less extreme environments [72,73]. We also emphasise that some physiological parameters (including bradycardia at great depths) show specific characteristics (high percentage of arrhythmias), at least in the trained bottlenose dolphin *T. truncatus* [74], indicating the persistence of ancestral terrestrial traits in cardiac functions that would be difficult to maintain during the routine deep foraging of beaked and sperm whales.

The reference tree used to map the CSBPSs (S1 Fig) was obtained from the analysis of the amino acid alignment. The length of the branches provides good evidence for the amount of

positive selection that occurred in the 13 proteins encoded in the mitochondrial genome. The branches connecting estuarine, freshwater and deep divers of Odontoceti are among the longest observed in the reference tree (S1 Fig). Indeed, they are significantly longer that those connecting other cetacean taxa ($p < 0.0005$; one-tailed Student's t-test, unequal sample sizes, and unequal variances). This finding supports the view that in these taxa, the proteins were subjected to positive selection, as shown previously for *P. macrocephalus* [75]. CSBPSs are much more abundant in taxa that are characterised by long branches (Fig 5; S1 Fig) than in other cetacean species ($p < 0.00005$; one-tailed Student's t-test, unequal sample sizes, and unequal variances). This match supports the view that not only did PCGs and their protein products experience positive selection, but tRNAs were subjected to an acceleration of the substitution process, which increased the number of CSBPSs. We suggest here that the challenging environments inhabited by estuarine, freshwater and deep water Odontoceti were responsible for at least part of the increased rate of base changes observed in the mitochondrial genomes of these taxa. Extreme environmental conditions have left their signature in the control region and in the coding genes of the mitochondrial genomes of high-altitude mammals [76]. Similarly, the mtDNA of the pika *Ochotona curzoniae* appears to harbour evidence of adaptation to cold and hypoxia [77]. In general, signatures of adaptive evolution have been found in the mitochondrial genomes of various mammals with specialised metabolic requirements [78].

Alternatively, it could be argued that the diverse numbers of CSBPSs observed in the various toothed whales are simply the result of a random substitution process linked to the different ages of the species. In this scenario, older species exhibit a higher number of CSBPSs because there was more time available for random substitution to occur.

To test this second hypothesis, which it is not necessarily an alternative to the environment-driven evolution described above, we compared the abundance of CSBPSs with the age of the taxa.

Different time estimations exist for the appearance and split of major phyletic lineages of Cetacea [12,15,55–58,62,63]. Dating is not consistent in different papers, and discrepancies exist (see the references cited above).

The currently most complete dating for the Cetacea clade is that provided by McGowen et al. [55]. According to these authors and considering only the dates relevant for the present paper, we have the following estimates (in brackets is the 95% interval range): (a) 24.21 MYA (15.83–31.93) for the split between *P. microcephalus* and *Kogia* genus; (b) 16.68 MYA (11.35–22.51) for the split between *I. geoffrensis* and *P. blainvillei*; c) 22.15 MYA (16.93–27.30) for the occurrence of the last common ancestor of *L. vexillifer* and *I. geoffrensis* + *P. blainvillei*; d) 32.43 MYA (27.92–37.07) for the appearance of the lineage leading to *Platanista* genus; e) 10.08 MYA (7.34–12.88) for the onset of Delphinidae; and f) 13.80 MYA (9.99–19.32) for the origin of Balaenopteridae + *E. robustus*. These estimations are more or less in agreement with the molecular dating based on complete mtDNA published by Hassanin et al. [15]. These authors provide the following estimates: a) 21.9 ± 3.6 MYA for the split between *P. macrocephalus* and *K. breviceps*; b) 14.0 ± 3.0 MYA for the split between *I. geoffrensis* and *P. blainvillei*; and c) 19.9 ± 3.2 MYA for the last common ancestor of *L. vexillifer* and *I. geoffrensis* + *P. blainvillei*.

Despite the variation in the absolute values of the estimates, the split between *P. macrocephalus* and *Kogia* occurred clearly before the separation between *I. geoffrensis* and *P. blainvillei*. Furthermore, *L. vexillifer* belongs to an older branch than the two species just mentioned. Finally, the differentiation of the lineage giving birth to the *Platanista* genus was a very early cladogenetic event. All of these taxa exhibit similar numbers of CSBPSs, and younger species may present more CSBPSs that do older species (*O. blainvillei* vs. *K. breviceps*). Similarly, even if we assigned to Balaenopteridae species the CSBPSs of the intermediated nodes present in the pathways connecting the root of this family with current taxa, thus spanning the whole 13.80

MYA of evolution, we still have much lower values than those of *P. blainvillei*. An analogous reasoning can be applied to Delphinidae.

As shown in the Results section, the substitution process in the stems of tRNAs is far from saturation. Thus, the observed distribution of CSBPSs cannot be explained in terms of pure random drift, even if random drift cannot be fully excluded and certainly plays/played some role in Cetacea tRNA evolution.

C. marginata was the only baleen whale with a high number of CSBPSs (Fig 5). This species is the sole living representative of a lineage thought to be extinct [79], and its physiology and habitat requirements are poorly known [80]. Thus, it is currently impossible to identify the main forces that shaped the evolution of mitochondrial tRNAs in this taxon.

Concluding remarks

During the transition from terrestrial to aquatic environments, the body plan and the physiology of Cetacea were extensively modified, and strong molecular signatures of these changes are becoming well documented in their nuclear genomes [81]. The results presented in this paper show that mitochondrial genomes harbour in their sequences evidence of the transition from terrestrial to aquatic environments and also permit differentiation among the different habitats currently inhabited by Cetacea.

The evolution of CSBPSs was not constant during the cladogenetic process that lead to current Cetacea and experienced two peaks of acceleration. The first peak occurred during the return to the water by the common ancestors of whales, dolphins and their relatives. The second one arose with the entering of some taxa into the more demanding environments represented by freshwater, estuarine and deep water habitats (Fig 5).

We outline here that the "extreme environment" hypothesis of the evolution of cetacean tRNAs represents the best interpretation of our data given the analyses performed in the present work. However, we do not claim that the evolution of the tRNAs was shaped only/mostly by harsh environmental conditions. The process was certainly influenced by other causes, including the genetic drift described above. Thus, further studies are necessary to improve our current understanding of the evolution of cetacean tRNAs.

Supporting Information

S1 Extended Results. File including more details on the analyses performed in this paper.
(PDF)

S1 Fig. The phylogeny of Cetartiodactyla. Maximum likelihood (-lnL = 66963.002109) phylogram depicting the phylogenetic relationships among the major clades of Cetartiodactyla. The tree was created by analysing the amino acid 94T-set (3716 positions) with the RAxML 7.4.2 program implemented in raxmlGUI 1.3.1. The evolutionary model was MTMAM + F + CAT. Thirteen partitions were applied: one for every protein. The numbers represent bootstrap values expressed in percent. Only bootstrap values $\geq$ 50% are provided for the nodes. The scale bar represents 0.05 substitutions/site.
(PDF)

S2 Fig. The mitochondrial genome of *Ziphius cavirostris*. The gene order is depicted and linearised starting from cox1. Genes encoded on the α-strand (right to left orientation) are underlined in green, while those encoded on the β-strand are underlined in red (left to right orientation). Gene nomenclature: *atp6* and *atp8*: *ATP synthase subunits 6* and *8*; *cob*: *apocytochrome b*; *cox1-3*: *cytochrome c oxidase sub-units 1–3*; *nad1-6* and *nad4L*: *NADH dehydrogenase subunits 1–6 and 4L*; *rrnS* and *rrnL*: small and large subunit ribosomal RNA (rRNA) genes;

and X: transfer RNA (tRNA) genes, where X is the one-letter abbreviation of the corresponding amino acid. In particular, L1 identifies the CTN codon family, L2 the TTR codon family, S1 the AGY codon family, and S2 the TCN codon family. CR, Control Region. OL, origin of duplication of the light strand. A black circle located between two adjacent genes denotes the presence of an intergenic spacer (in white is the number of nucleotides forming the spacer). A white circle located between two adjacent genes denotes the presence of an overlapping segment (in red is the number of nucleotides forming the segment). Isp, intergenic spacer; start, start of the gene; end, end of the gene; size, size of the gene. For protein-coding genes, the start codon is provided in cyan and the stop codon in red (with incomplete stop codons written in parentheses). The anticodon is provided for every tRNA (e.g., tga for *trnS2*).
(PDF)

S3 Fig. Secondary structure of Cetacea tRNA and level of conservation (*trnA-trnK*). pDis, p-Distance calculated for each pairwise-comparison orthologous tRNAs. MLdis, maximum composite likelihood distance calculated for every pairwise-comparison. DIF (MLdis–pDis), the difference between MLdis and pDis. The average values and the standard deviation are provided for both pDis and DIF. The values were computed for each set of orthologous tRNAs.
(PDF)

S4 Fig. Secondary structure of Cetacea tRNA and level of conservation (*trnL1-trnV*). pDis, p-Distance calculated for each pairwise-comparison orthologous tRNAs. MLdis, maximum composite likelihood distance calculated for every pairwise-comparison. DIF (MLdis–pDis), the difference between MLdis and pDis. The average values and the standard deviation are provided for both pDis and DIF. The values were computed for each set of orthologous tRNAs.
(PDF)

S5 Fig. AT-skew vs. A+T% and GC-skew vs. G+C% in the 94T-set mtDNAs. The values were calculated on the α-strand of the full-length mtDNA genomes. The X axis provides the skew values, while the Y axis provides the A+T% and G+C% values.
(PDF)

S6 Fig. AT-skew vs. A+T% (A), and GC-skew vs. G+C% (B) in Cetacea mtDNAs. The values were calculated on the α-strand of the full-length mtDNA genomes. The X axis provides the AT- and GC-skew values, while the Y axis provides the A+T% and G+C% values. Species with a placement that is difficult to identify in the main plots are depicted in the frames.
(PDF)

S7 Fig. Mapping of FCBCs on the Cetacea phylogenetic tree. FCBC, fully compensatory base change; SPIC, stem position involved in base change. The tRNA and the stem pair involved in FCBCs are mapped on the corresponding nodes of the reference phylogenetic tree. The tRNAs are depicted with the single-letter IUPAC code used for the corresponding amino acid. In particular, L1 identifies the CTN codon family, L2 the TTR codon family, S1 the AGY codon family, and S2 the TCN codon family. The stem pair involved in FCBC is provided in superscript. The asterisk, associated with some FCBCs indicates that these FCBCs were subjected to successive changes in one/some of the taxa located downstream of the considered node.
(PDF)

S8 Fig. Mapping of HFBCs on the Cetacea phylogenetic tree. HFBC, hemi-compensatory base change; SPIC, stem position involved in base change. The tRNA and the SPICs involved in HFBCs are mapped on the corresponding nodes of the reference phylogenetic tree. The tRNAs are depicted with the single-letter IUPAC code used for the corresponding amino acid. In particular, L1 identifies the CTN codon family, L2 the TTR codon family, S1 the AGY

codon family, and S2 the TCN codon family. The SPIC involved in HFBC is provided in superscript. The asterisk associated with some HFBCs indicates that these HFBCs were subjected to successive changes in one/some of the taxa located downstream of the considered node. A SPIC located on the 5' side of a stem-pair is marked in orange, while a SPIC placed on the 3' end of a pair is purple.
(PDF)

S9 Fig. Mapping of Mismatches on the Cetacea phylogenetic tree. Mismatch, mismatch in a base pair of a stem; SPIC, stem position involved in base change. The tRNAs and SPICs involved in mismatches are mapped on the corresponding nodes of the reference phylogenetic tree. The tRNAs are depicted with the single-letter IUPAC code used for the corresponding amino acid. In particular, L1 identifies the CTN codon family, L2 the TTR codon family, S1 the AGY codon family, and S2 the TCN codon family. The SPIC involved in a mismatch is provided in superscript. The asterisk associated with some mismatches indicates that these mismatches were subjected to successive changes in one/some of the taxa located downstream of the considered node. A SPIC located on the 5' side of a stem-pair is marked in orange, and a SPIC placed on the 3' end of a pair is purple.
(PDF)

S1 Table. List of taxa, accession numbers in GenBank and references.
(PDF)

S2 Table. Intraspecific CSBPSs identified in some Cetacea.
(PDF)

S1 tRNA Multiple Alignments. Multiple alignments of orthologous tRNAs.
(PDF)

Acknowledgments

The Authors acknowledge the help of the Mediterranean Marine Mammal Tissue Bank of the University of Padova, which supplied the tissue samples, and Dr. Maristella Giurisato for her technical and scientific contributions. We thank Prof. Piero Fariselli (University of Padova) for helpful advice regarding statistical tests. We express our thanks to two anonymous reviewers who provided constructive criticisms to an early version of the manuscript. This work was supported by a grant to AP (University of Padova ex-60% 2012) and by a grant to EN (University of Padova ex-60% 2012).

Author Contributions

Conceived and designed the experiments: AP EN. Performed the experiments: SM EN. Analyzed the data: EN. Wrote the paper: SM AP TP BC EN.

References

1. Cameron SL, Dowton M, Castro LR, Ruberu K, Whiting MF, Austin AD, et al. Mitochondrial genome organization and phylogeny of two vespid wasps. Genome 2008; 51:800–808. doi: 10.1139/G08-066 PMID: 18923531
2. Thewissen JGM, Cooper LN, Clementz MT, Bajpai S, Tiwari BN. Whales originated from aquatic artiodactyls in the Eocene epoch of India. Nature 2007; 450: 1190–1194. PMID: 18097400
3. Uhen MD. Evolution of marine mammals: back to the sea after 300 million years. Anat Rec (Hoboken) 2007; 290: 514–522.
4. Perrin WF. World Cetacea Database. 2015. Available: http://www.marinespecies.org/cetacea.

5. Arnason U, Gullberg A. Comparison between the complete mtDNA sequences of the blue and the fin whale, two species that can hybridize in nature. J Mol Evol. 1993; 37: 312–322. PMID: 8308901
6. Valverde JR, Marco R, Garesse R. A conserved heptamer motif for ribosomal RNA transcription termination in animal mitochondria. Proc Natl Acad Sci U.S.A. 1994; 91: 5368–5371. PMID: 7515499
7. Arnason U, Gullberg A, Gretarsdottir S, Ursing B, Janke A. The mitochondrial genome of the sperm whale and a new molecular reference for estimating eutherian divergence dates. J Mol Evol. 2000; 50: 569–578. PMID: 10835487
8. Arnason U, Gullberg A, Janke A. Mitogenomic analyses provide new insights into cetacean origin and evolution. Gene 2004; 333: 27–34. PMID: 15177677
9. Sasaki T, Nikaido M, Hamilton H, Goto M, Kato H, Kanda N, et al. Mitochondrial phylogenetics and evolution of mysticete whales. Syst Biol. 2005; 54: 77–90. PMID: 15805012
10. Yan J, Zhou K, Yang G. Molecular phylogenetics of 'river dolphins' and the baiji mitochondrial genome. Mol Phylogenet Evol. 2005; 37: 743–750. PMID: 16046150
11. Sasaki T, Nikaido M, Wada S, Yamada TK, Cao Y, Hasegawa M, et al. *Balaenoptera omurai* is a newly discovered baleen whale that represents an ancient evolutionary lineage. Mol Phylogenet Evol. 2006; 41: 40–52. PMID: 16843687
12. Xiong Y, Brandley MC, Xu S, Zhou K, Yang G. Seven new dolphin mitochondrial genomes and a time-calibrated phylogeny of whales. BMC Evol Biol. 2009; 9: 20. doi: 10.1186/1471-2148-9-20 PMID: 19166626
13. Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, et al. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. Genome Res. 2010; 20: 908–916. doi: 10.1101/gr.102954.109 PMID: 20413674
14. Vilstrup JT, Ho SYW, Foote AD, Morin PA, Kreb D, Krützen M, et al. Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. BMC Evol Biol. 2011; 11: 65. doi: 10.1186/1471-2148-11-65 PMID: 21392378
15. Hassanin A, Delsuc F, Ropiquet A, Hammer C, Jansen van Vuuren B, Matthee C, et al. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. C R Biol. 2012; 335: 32–50. doi: 10.1016/j.crvi.2011.11.002 PMID: 22226162
16. Moura AE, Nielsen SC, Vilstrup JT, Moreno-Mayar JV, Gilbert MT, Gray HW, et al. Recent diversification of a marine genus (*Tursiops* spp.) tracks habitat preference and environmental change. Syst Biol. 2013; 62: 865–877. doi: 10.1093/sysbio/syt051 PMID: 23929779
17. Thompson KF, Patel S, Williams L, Tsai P, Constantine R, Baker CS, et al. High coverage of the complete mitochondrial genome of the rare Gray's beaked whale (*Mesoplodon grayi*) using Illumina next generation sequencing. Mitochondrial DNA 2014. In press doi: 10.3109/19401736.2013.878908
18. Boore JL. Animal mitochondrial genomes. Nucleic Acids Res. 1999; 27: 1767–1780. PMID: 10101183
19. Suzuki T, Nagao A, Suzuki T. Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. Annu Rev Genet. 2011; 45: 299–329. doi: 10.1146/annurev-genet-110410-132531 PMID: 21910628
20. Sprinzl M, Horn C, Brown M, Ioudovitch A, Steinberg S. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. 1998; 26: 148–153. PMID: 9399820
21. Cannone J, Subramanian S, Schnare MN, Collet JR, D'Souza LM, Du Y, et al. The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics 2002; 3: 2. PMID: 11869452
22. Coleman AW. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends Genet. 2003; 19: 370–375. PMID: 12850441
23. Robertus JD, Ladner JE, Finch JT, Rhodes D, Brown RS, Clark BF, et al. Structure of yeast phenylalanine tRNA at 3 A resolution. Nature 1974; 250: 546–551. PMID: 4602655
24. Zuker M, Jaeger JA, Turner DH. A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures determined by phylogenetic comparison. Nucleic Acids Res. 1991; 19: 2707–2714. PMID: 1710343
25. Patwary MU, Kenchington EL, Bird CJ, Zouros E. The use of random amplified polymorphic DNA markers in genetic studies of the sea scallop *Placopecten magellanicus* (Gmelin, 1791). J Shellfish Res. 1994; 13: 547–553.
26. Cabria MT, Rubines J, Gómez-Moliner B, Zardoya R. On the phylogenetic position of a rare Iberian endemic mammal, the Pyrenean desman (*Galemys pyrenaicus*). Gene 2006; 375:1–13. PMID: 16624502

27. Negrisolo E, Babbucci M, Patarnello T. The mitochondrial genome of the ascalaphid owlfly *Libelloides macaronius* and comparative evolutionary mitochondriomics of neuropterid insects. BMC Genomics 2011; 12: 221. doi: 10.1186/1471-2164-12-221 PMID: 21569260

28. Babbucci M, Basso A, Scupola A, Patarnello T, Negrisolo E. Is it an ant or a butterfly? Convergent evolution in the mitochondrial gene order of Hymenoptera and Lepidoptera. Genome Biol Evol. 2014; 6: 3326–3343. doi: 10.1093/gbe/evu265 PMID: 25480682

29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215: 403–410. PMID: 2231712

30. Tatusova TA, Madden TL. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett. 1999; 174: 247–250. PMID: 10339815

31. Wolstenholme DR. Animal mitochondrial DNA: structure and evolution. Int Rev Cytol. 1992; 141: 173–216. PMID: 1452431

32. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22: 4673–4680. PMID: 7984417

33. Smith AE, Marcker KA. N-formylmethionyl transfer RNA in mitochondria from yeast and rat liver. J Mol Biol. 1968; 38: 241–243. PMID: 5760639

34. Fearnley IM, Walker JE. Initiation codons in mammalian mitochondria: differences in genetic code in the organelle. Biochemistry 1987; 26: 8247–8251. PMID: 2964865

35. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997; 25: 955–964. PMID: 9023104

36. Helm M, Brulé H, Friede D, Giegé R, Pütz, Florentz C. Search for characteristic structural features of mammalian mitochondrial tRNAs. RNA 2000; 6: 1356–1379. PMID: 11073213

37. LeDuc RG, Robertson KM, Pitman RL. Mitochondrial sequence divergence among Antarctic killer whale ecotypes is consistent with multiple species. Biol Lett. 2008; 4: 426–429. doi: 10.1098/rsbl.2008.0168 PMID: 18524738

38. Foote AD, Morin PA, Durban JW, Willerslev E, Orlando L, Gilbert MT. Out of the Pacific and back again: insights into the matrilineal history of Pacific killer whale ecotypes. PLoS ONE 2011; 6: e24980. doi: 10.1371/journal.pone.0024980 PMID: 21949818

39. de Bruyn PJN, Tosh CA, Terauds A. Killer whale ecotypes: is there a global model? Biol Rev Camb Philos Soc. 2013; 88: 62–80. doi: 10.1111/j.1469-185X.2012.00239.x PMID: 22882545

40. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 2010; 38: W7–W13. doi: 10.1093/nar/gkq291 PMID: 20435676

41. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transformation. Nucleic Acids Res. 2002; 30: 3059–3066. PMID: 12136088

42. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 2005; 33: 511–518. PMID: 15661851

43. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000; 17: 540–552. PMID: 10742046

44. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011; 28: 2731–2739. doi: 10.1093/molbev/msr121 PMID: 21546353

45. Perna NT, Kocher TD. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol. 1995; 41: 353–358. PMID: 7563121

46. Negrisolo E, Minelli A, Valle G. The mitochondrial genome of the house centipede *Scutigera* and myriapod monophyly vs. paraphyly. Mol Biol Evol. 2004; 21: 770–780. PMID: 14963096

47. Felsenstein J. Inferring phylogenies. Sunderland (MA): Sinauer Associates; 2004.

48. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006; 22: 2688–2690. PMID: 16928733

49. Silvestro D, Michalak I. raxmlGUI: a graphical front-end for RAxML. Organ Div Evol. 2011; doi: 10.1007/s13127-011-0056-0

50. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 1985; 39: 783–791.

51. Lanave C, Preparata G, Saccone C, Serio G. A new method for calculating evolutionary substitution rates. J Mol Evol. 1984; 20: 86–93. PMID: 6429346

52. Lartillot N, Philippe H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol. 2004; 21: 1095–1109. PMID: 15014145

53. Yang Z, Nielsen R, Hasegawa M. Models of amino acid substitution and applications to mitochondrial protein evolution. Mol Biol Evol. 1998; 15: 1600–1611. PMID: 9866196

54. Agnarsson I, May-Collado LJ. The phylogeny of Cetartiodactyla: the importance of dense taxon sampling, missing data, and the remarkable promise of cytochrome b to provide reliable species-level phylogenies. Mol Phylogenet Evol. 2008; 48: 964–985. doi: 10.1016/j.ympev.2008.05.046 PMID: 18590827

55. McGowen MR, Spaulding M, Clark C, Gatesy J. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. Mol Phylogenet Evol. 2009; 53: 891–906. doi: 10.1016/j.ympev.2009.08.018 PMID: 19699809

56. Chen Z, Xu S, Zhou K, Yang G. Whale phylogeny and rapid radiation events revealed using novel retroposed elements and their flanking sequences. BMC Evol Biol. 2011; 11: 314. doi: 10.1186/1471-2148-11-314 PMID: 22029548

57. Duchêne S., Archer FI, Vilstrup J, Caballero S, Morin PA. Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. PLoS ONE 2011; 6: e27138. doi: 10.1371/journal.pone.0027138 PMID: 22073275

58. Geisler JH, McGowen MR, Yang G, Gatesy J. A supermatrix analysis of genomic, morphological, and paleontological data from crown Cetacea. BMC Evol Biol. 2011; 11: 112. doi: 10.1186/1471-2148-11-112 PMID: 21518443

59. McGowen MR. Toward the resolution of an explosive radiation-a multilocus phylogeny of oceanic dolphins (Delphinidae). Mol Phylogenet Evol. 2011; 60: 345–357. doi: 10.1016/j.ympev.2011.05.003 PMID: 21600295

60. Zhou X, Xu S, Yang Y, Zhou K, Yang G. Phylogenomic analyses and improved resolution of Cetartiodactyla. Mol Phylogenet Evol. 2011; 61: 255–264. doi: 10.1016/j.ympev.2011.02.009 PMID: 21315162

61. Dornburg A, Brandley MC, McGowen MR, Near TJ. 2Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). Mol Biol Evol. 2012; 9: 721–736.

62. Alexander A, Steel D, Slikas B, Hoekzema K, Carraher C, Parks M, Cronn R, Baker CS. Low diversity in the mitogenome of sperm whales revealed by next-generation sequencing. Genome Biol Evol. 2013; 5: 113–129. doi: 10.1093/gbe/evs126 PMID: 23254394

63. Gatesy J, Geisler JH, Chang J, Buell C, Berta A, Meredith RW, et al. A phylogenetic blueprint for a modern whale. Mol Phylogenet Evol. 2013; 66: 479–506. doi: 10.1016/j.ympev.2012.10.012 PMID: 23103570

64. Marx FG, Fordyce RE. Baleen boom and burst: a synthesis of mysticete phylogeny, diversity and disparity. R Soc Open Sci. 2015; 2: 140434. doi: 10.1098/rsos.140434 PMID: 26064636

65. Maddison WP, Maddison DR. Mesquite: a modular system for evolutionary analysis. Version 3.02 2015. Available: http://mesquiteproject.org.

66. Meer MV, Kondrashov AS, Artzy-Randrup Y, Kondrashov FA. Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. Nature 2010; 464: 279–283. doi: 10.1038/nature08691 PMID: 20182427

67. Yim HS, Cho YS, Guang X, Kang SG, Jeong JY, Cha SS, et al. Minke whale genome and aquatic adaptation in cetaceans. Nat Genet. 2014; 46: 88–92. doi: 10.1038/ng.2835 PMID: 24270359

68. Nery MF, Arroyoi JI, Opazo JC. Increased rate of hair keratin gene loss in the cetacean lineage. BMC Genomics 2014; 15: 869. doi: 10.1186/1471-2164-15-869 PMID: 25287022

69. Ai WM, Chen SB, Chen X, Shen XJ, Shen YY. Parallel evolution of IDH2 gene in cetaceans, primates and bats. FEBS Lett. 2014; 588: 450–454. doi: 10.1016/j.febslet.2013.12.005 PMID: 24374336

70. Wang Z, Yuan L, Rossiter SJ, Zuo X, Ru B, Zhong H, et al. Adaptive evolution of 5'HoxD genes in the origin and diversification of the cetacean flipper. Mol Biol Evol. 2009; 26: 613–622. doi: 10.1093/molbev/msn282 PMID: 19074008

71. McGowen MR, Montgomery SH, Clark C, Gatesy J. Phylogeny and adaptive evolution of the brain-development gene microcephalin (MCPH1) in cetaceans, BMC Evol Biol. 2011, 11: 98. doi: 10.1186/1471-2148-11-98 PMID: 21492470

72. Ortiz RM. Osmoregulation in marine mammals. J Exp Biol. 2001; 204: 1831–1844. PMID: 11441026

73. Tyack PL, Johnson M, Aguilar Soto N, Sturlese A, Madsen PT. Extreme diving of beaked whales. J Exp Biol. 2006; 209: 4238–4253. PMID: 17050839

74. Williams TM, Fuiman LA, Kendall T, Berry P, Richter B, Noren SR, et al. Exercise at depth alters bradycardia and incidence of cardiac anomalies in deep-diving marine mammals. Nat Commun. 2014; 6: 6055.

75. Foote AD, Morin PA, Durban JW, Pitman RL, Wade P, Willerslev E, et al. Positive selection on the killer whale mitogenome. Biol Lett. 2011; 7:116–118. doi: 10.1098/rsbl.2010.0638 PMID: 20810427

76. Hassanin A, Ropiquet A, Couloux A, Cruaud C. Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). J Mol Evol. 2009; 68: 293–310. doi: 10.1007/s00239-009-9208-7 PMID: 19294454

77. Luo Y, Gao W, Gao Y, Tang S, Huang Q, Tan X, et al. Mitochondrial genome analysis of *Ochotona curzoniae* and implication of cytochrome c oxidase in hypoxic adaptation. Mitochondrion 2008; 8: 352–357. doi: 10.1016/j.mito.2008.07.005 PMID: 18722554

78. da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A. The adaptive evolution of the mammalian mitochondrial genome. BMC Genomics 2008; 9: 119. doi: 10.1186/1471-2164-9-119 PMID: 18318906

79. Fordyce RE, Marx FG. The pygmy right whale *Caperea marginata*: the last of the cetotheres. Proc Royal Soc B Biol Sci. 2013; 280: 20122645 (doi: 10.1098/rspb.2012.2645).

80. Best PB. Family Neobalaenidae. Pigmy right whale. *Caperea marginata*. In Best P, editor. Whales and dolphins of the Southern African subregion. Cape Town (SA): Cambridge University Press; 2007. pp. 36–40.

81. McGowen MR, Gatesy J, Wildman DE. Molecular evolution tracks macroevolutionary transitions in Cetacea. Trend Ecol Evol. 2014; 29: 336–346.

S1 Supplementary Extended Results

The mitochondrial genome of *Ziphius cavirostris*

The mtDNA of a specimen of *Z. cavirostris,* sequenced for this paper, is briefly described here. The new mitochondrial genome was 16,352 bp long. This value was very close to the average value obtained for the dataset analysed in the present work (16,436 ± 124). The *Z. cavirostris* genome contained the 37 genes almost universally found in animal mtDNAs i.e., 13 PCGs, two ribosomal rRNAs and 22 tRNAs. The gene order was typical for vertebrate mtDNAs (Fig 1), with 28 genes encoded on the α-strand and nine present on the opposite β-strand. Most of the PCGs started with ATG and ended with TAA or the incomplete stop codons TA(a) and T(aa). The genes on the same/opposite strand overlapped, were contiguous or were separated by intergenic spacers encompassing a variable number of nucleotides (S2 Fig). The mtDNA sequence of *Z. cavirostris* is available in EBI/GenBank under accession number LN997430.

Occurrence of CSBPSs in the tRNAs of Cetacea

The computation of values of the p-distance and (maximum composite likelihood distance – p-distance) difference allowed the level of conservation and the possible underestimation of the substitution patterns in Cetacea tRNAs (S3 and S4 Figs) to be determined. The most conserved tRNA was *trnG* (average-pDis = 0.049 ± 0.026) and the most variable was *trnH* (average-pDis = 0.134 ± 0.059). The minimum DIFF value (average-DIF = 0.000 ± 0.002) and the maximum value (average-DIF = 0.027 ± 0.019) were observed respectively in *trnE* and *trnH*. The (maximum composite likelihood distance – p-distance) difference values demonstrated that the observed CSBPSs did not grossly underestimate the true number of CSBPSs in the tRNAs.

A total of 603 CSBPSs (136 FCBCs, 320 HCBCs, 147 mismatches) were identified in the tRNAs of Cetacea (Fig 2; S10 tRNAs multiple alignments).

Most of FCBCs were of type I (131). These FCBCs were split in 36 A–T vs. G–C, 49 G–C vs. A–T, 24 C–G vs. T–A, 22 T–A vs C–G. The distribution on α-/β-strand tRNAs of type I FCBCs was 35/1, 38/11, 19/5, and 22/2. Four FCBCs were of type II. T–A vs. A–T occurred on the 5an-pair of *trnH*. G–T vs. T–A was found on the 7ac-pair of *trnQ*. C–G vs. A–T was identified on the 2tp-pair of *trnE*. A–T vs. T–A was detected on the 2tp-pair of *trnN* (Fig 2; S10 tRNAs multiple alignments). This latter case showed the complete pathway of the process leading to a type II FCBC. Indeed the ougroups had the T–A pair, most of the ingroup taxa exhibited the mismatch T|T, and *K. brevices* and *Mesoplodon grayi* (see below for this latter species) exhibited the A–T pair. Finally, a peculiar FCBC

occurred in *trnF* (i.e. G•T vs. A|C), where a double change of bases generated a correct pairing against a mismatch.

The number of FCBCs, detected in a single tRNA, ranged from 1 (*trnC, trnE*, *trnI*, *trnN, trnS2*) to 18 *(trnH*). Ten or more FCBCs were found only in the α-strand encoded *trnF*, *trnH*, *trnK*, *trnR*, *trnS1* (Figs 1 and 2). Except for *trnY* (8 FCBCs), the β-strand tRNAs exhibited four or less FCBCs. The bases alternating on the type I FCBCs followed three patterns (a-c). (a) Only purines alternated at the 5' end of the pair and solely pyrimidines occurred at the 3'end (*trnD*, *trnG*, *trnI*, *trnL2* and *trnW*). (b) Both purine/pyrimidine bases occurred in the substitution at 5' and 3'ends of the pair (*trnA*, *trnF*, *trnH*, *trnK*, *trnL1*, *trnP*, *trnR*, *trnS1, trnT, trnV, trnY*) with a variable prevalence of the first/second type of base. (c) Only pyrimidines alternated at the 5'end and only purines occurred at the 3'end of the pair (*trnC*, *trnM*, *trnS2*) (S10 tRNAs multiple alignments).

The 320 HCBCs were distributed mainly in four symmetrical types a-d: (a) [5' (A vs. G), 3' (T)] (43) and its counterpart [5' (T), 3' (A vs. G)] (16); (b) [5'(C vs. T), 3' (G)] (6) and its opposite [5' (G), 3' (C vs. T)] (26); (c) [5' (G vs. A), 3' (T)] (67) and the opposite [5' (T), 3' (G vs. A)] (60); (d) [5' (T vs. C), 3' (G)] (23) and its counterpart [5' (G), 3' (T vs. C)] (33). Abundant was also the type (31) [5' (A), 3' (T vs. C)]. The α-strand tRNAs showed 144 HCBCs, while the β-strand tRNAs exhibited 176 HCBCs (Fig 2).

The HCBCs had a distribution β-strand biased. The maximum number of HCBCs occurred in *trnE* (35). Numerous HCBCs (≥23) were present in *trnA*, *trnP*, *trnQ*, *trnS2*, and *trnY*. The *trnC* (11 HCBCs) and more markedly *trnN* (0 HCBC) were exceptions. The HCBCs distribution was variable in the α-strand tRNAs with values ranging from 17 (*trnH*) to 2 (*trnM*, *trnV*).

The 147 mismatches belonged to 35 different types (Fig 2). Only the most abundant are described in details below. The symmetrical [5' (A vs. G), 3'(C)] and [5'(C), 3' (A vs. G)] mismatches occurred respectively 25 and 6 times. The [5' (C vs. T), 3' (A)] and [5' (A), 3' (C vs. T)] mismatches were found respectively 23 and 34 times. The remaining 31 types occurred 59 times and accounted for the 40.14% of the whole mismatches set. The distribution of mismatches was uneven and α-strand biased. Indeed ten α-strand tRNAs (i.e., *trnF*, *trnH*, *trnK*, *trnL2*, *trnM*, *trnR*, *trnS1*, *trnT, trnV* and *trnW*) accounted for most (131, 89.16%) of observed mismatches. Conversely, only seven mismatches occurred in the eight tRNAs of the β-strand.

Patterns of CSBPSs distribution in the Cetacea tRNAs

The distribution of CSBPSs was very variable in the tRNAs (Fig 2). Eight tRNAs of the α-strand (*trnF*, *trnH*, *trnK*, *trnL2*, *trnR*, *trnS1*, *trnT* and *trnW*) exhibited a linked distribution of FCBCs, HCBCs, and mismatches with all types of CSBPS ≥ 7. A second pattern, mainly observed in β-strand

tRNAs, implied a high number of HCBCs coupled with a low number of FCBCs and mismatches (*trnA*, *trnC*, *trnE*, *trnG*, *trnP*, *trnQ*, *trnS2*). The *trnD*, *trnL1*, and *trnY* exhibited a low number of mismatches coupled with a moderate number of FCBCs and a higher number of HCBCs. A low numbers of base changes characterised *trnI*, and *trnN*. Few FCBCs and HCBCs and good number of mismatches occurred in *trnM*. Finally, *trnV* exhibited a high number of mismatches coupled with a good number of FCBCs and a low number of HCBCs (Fig 2; S10 tRNAs multiple alignments).

Factors influencing the occurrence and type of CSBPSs

The total number of codons encoded by the cetacean mtDNAs was homogenous (average 3777.47 ± 1.23). The minimum (3775) and maximum values (3781) were observed respectively in *Orcaella brevirostris* and *P. minor*. The number of codons for each codon family was also homogenous. The most abundant family was Leu1 (codons per thousand codons = 133.10 ± 3.72). The less abundant family was Cys (codons per thousand codons = 6.07 ± 0.22) (Fig 2).

FCBCs, HCBCs, and mismatches were simultaneously abundant in α-strand tRNAs exhibiting a broad range of codons per thousand codons (*trnH*, *trnK*, *trnR*, *trnS1*, *trnL2*, *trnF*, *trnT*) (Fig 2). Mismatches were very low/absent in all the β-strand tRNAs. Mismatches were very low in *trnL1* (1) and *trnI* (2), both encoded on α-strand, representing the first and third most abundant family. The smallest number of CSBPSs occurred in *trnN* (0 FCBC; 0 HCBC; 1 mismatch), one of the less abundant families (codons per thousand codons = 25.35 ± 0.50). On the opposite *trnI*, one of the most abundant tRNAs, ranked as the second for the minimum number of FCBCs (1), HCBCs (3) and mismatches (2). In general, there was not a simple pattern linking codon family abundances and richness of FCBCs, HCBCs, and mismatches (Fig 2).

The bases contents, the AT-/GC-skews were computed for the tRNAs stems and compared with those of the encoding mtDNAs to test their effects on CSBPSs distribution (Fig 2; S5 and S6 Figs).

The tRNAs of β-strand exhibited negative/null AT-skew, and positive/null GC-skews values (Fig 2). The α-strand tRNAs showed a more composite pattern. The *trnS1* had always positive AT-skew values. In other tRNAs the AT-skews were negative/null (e.g., *trnG*), or varied from negative to positive values (e.g. *trnF*) within a more or less broad range. The GC-skew values were positive in *trnL1*, and positive/null in *trnG*. In other tRNAs the GC-skew values varied from negative to positive. The A+T contents were very variable (Fig 2).

The majority of tRNAs exhibited always A+T content ≥ 50%. However, *trnC*, *trnF*, *trnL2*, *trnP*, *trnR*, *trnS2*, *trnW*, and *trnY* exhibited some sequences with A+T < 50%. The A+T content was always < 50% in *trnM* stems.

On α-strand: (a) the tRNAs (i.e., *trnD*, *trnI*) with high A+T content (≥ 62.50%) and limited variation of AT-skews (± 0.04) and GC-skews (± 0.091) were linked to a small number of FCBCs; (b) a low A+T content (≤ 47.62%) combined with small AT-skew values (-0,176 – -0.059), and small GC-skew values (-0.040 – 0.043) was associated to a small number of FCBCs in *trnM*; (c) a high variation of A+T content coupled to a broad range variation of AT-, and GC-skews was linked to a high number of CSBPSs (e.g., *trnF*, *trnH*, *trnL2*, *trnR, trnT*). On β-strand a combination of always negative AT-skews and always positive GC-skews coupled with variable A+T content was linked to a high number of HCBCs, and low number of both FCBCs and mismatches (e.g., *trnA*, *trnE*, *trnP*, *trnQ*). An invariable A+T content, limited variation of AT-skews, and constant GC-skews characterized *trnN*, that exhibited one FCBC, one mismatch and none HCBC.

Globally, the occurrence of CSBPSs in different tRNAs was influenced by the combined action of the base content variation, and asymmetrical compositional biases of the stems that in several cases were opposite to the values computed for the strand encoding the analysed tRNAs. Particularly the range of variation and the fluctuation of bases content, AT-,and GC-skews had a major impact on the type and abundance of CSBPs (e.g., *trnH* vs. *trnN*) (Fig 2; S5 and S6 Figs).

Finally, the abundance of FCBCs, HCBCs, and mismatches did not appear to be linked to the genomic placement of different tRNAs (Figs 1 and 2). A couple of examples support this statement. *TrnA* and *trnN,* both on β-strand and adjacent, exhibited very different behaviours. Conversely *trnR* and *trnT*, both on α-strand and well separated, had very similar patterns.

The stem-positions associated to CSBPSs

The stem positions involved in base changes (hereafter named SPICs) were mapped and analysed in the different tRNAs (Figs 3 and 4). One, two and even all three types of substitution were observed in the same SPIC (e.g., *trnF*, Fig 3). The number of SPICs was very variable within the 22 tRNAs (Fig 3). The smallest number (2) of SPICs (5.00 % of the 40 stem-positions) occurred in *trnN*. At the opposite, *trnH* presented 23 SPICs (54.76% of the 42 stem-positions) (Fig 3). Due to the heterogeneity of the substitution patterns, a perfect correspondence did not exist between the percentage of SPICs and global percentage of CSBPSs occurring in a single tRNA. Thus, *trnH*, exhibiting the maximum number of SPICs, hosted 7.30% of the total 603 CSBPSs. Conversely, *trnR,* having a lower percentage of SPICs (47.62%) hosted 9.45 % of total CSBPSs. Similarly, *trnQ* with only 11 SPICs (26.19%) contained 5.47% of total CSBPSs, a value observed in tRNAs with SPICs percentages ≥ 35% (e.g., *trnY*). Despite these vagaries, the percentage of SPICs was in good agreement with the global percentage of CSBPSs. The percentages of SPICs and FCBCs exhibited a similar behaviour. Much more discrepancies existed among the percentage of SPICs and global

percentages of HCBCs, and mismatches. A couple of examples corroborate this statement. The SPICs percentage for *trnE* was 35.71 % while the HCBCs percentage was 10.94%, more than twice the percentage of HCBCs (5.31%) observed in *trnH*, that possessed the highest percentage of SPICs (see above). Similarly, *trnH* and *trnM* presented comparable percentages (6.12% vs. 5.44%) of mismatches but very different percentages of SPICs (54.76% vs. 26.19%) (Fig 3). Every tRNA exhibited a unique pattern of SPICs and associated types of FCBCs, HCBCs, mismatches.

The distributions of SPICs and CSBPSs were summarised in Fig 4. The occurrence of FCBCs was very variable in the pairs. None tRNA presented a FCBC in the 2ac-, and 4dh-pairs (Fig 4a). Conversely the 4ac-pair (7 tRNAs), the 2an-pair (7 tRNAs), and 2an-pair (8 tRNAs) were hot spots for the presence of FCBCs. The acceptor stem (3-6ac-pairs), the anticodon stem (1-2,5an-pairs) and the TΨC stem (2-4tp-pairs) contained most of SPICs associated to FCBCs. The DHU stem had a very limited number of SPICs associated to FCBCs.

The SPICs associated to HCBCs were variably distributed in the different tRNA. However, some hot spots emerged. The 5'end of dh1-pair hosted at least an HCBCs in nine tRNAs. On the opposite, the 3'end of 2dh-pair never exhibited an HCBC and the same was true for the 5'end of the 5tp-pair. The 5'-end and 3'-end could behave differently in the same pair. Thus, the 5'end of 3tp-pair hosted an HCBC in seven different tRNA while the 3'end presented an HCBC only in two tRNAs. No HCBC occurred in 3'-end of 2dh-pair.

The SPICs associated to mismatches were more abundant on acceptor and TΨC stems. The anticodon stem presented several SPICs associated to mismatches. Very a few SPICs hosting mismatches occurred in the DHU stem. Mismatches were never detected in some positions (e.g., 5'end of 1ac-pair, 4dh-pair, 1tp-pair) (Fig 4a). When the global percentage of FCBCs, HCBCs, mismatches occurring at the different SPICs was evaluated, the patterns that emerged largely mirrored the abundance of SPICs just described above (Fig 4b).

The occurrence of HCBCs exhibited an evident 5'end or 3'end distributional bias in the pairs of some tRNAs. In *trnG* 11 of the 12 HCBCs occurred at the 5'end of the involved pairs (Fig 3; S8 Fig). Likewise, in *trnH* 15 of the 17 HCBCs were located in the 3'end of the pairs (Fig 3; S8 Fig). Also, *trnF*, *trnK*, *trnR*, *trnS1*, *trnT* exhibited HCBCs distributional biases. The distribution of mismatches exhibited an evident 5'end or 3'end bias in *trnF*, *trnG*, *trnR*, *trnS1*, and *trnW* (Fig 3; S9 Fig).

Finally the known distribution of positions in the stems, where occur posttranscriptional modifications, was mapped and compared with the SPICs behaviour (Fig 4). A simple pattern linking these positions with CSBPSs/ SPICs was not identified.

Phylogenetic distribution of CSBPSs

The distribution of CSBPSs along the reference tree is summarized in Fig 5 while the full details are provided in the S7-S9 Figs.

The 70.32% of the CSBPSs were associated to living species of Cetacea while the remaining 29.68% was divided among the internal nodes of the tree. In living species, the percentage of FCBCs was 66.91%, that of HCBCs was 72.19%, and that of mismatches was 69.39% (Fig 5).

The four FCBCs of type II identified in Cetacea (see above) had a variable taxonomic distribution. Two were restricted to a single species i.e. *K. breviceps* (*trnN*), and *I. geoffrensis* (*trnE*) (S10 tRNAs multiple alignments). The FCBC occurring in *trnH* appeared at the onset of Cetacea. During the cladogenetic process, successive FCBCs of type I, HCBCs and a mismatch substituted this CSBPS in some cetacean species (S7-S9 Figs, S10 tRNAs multiple alignments, *trnH*). Finally, the FCBC of type II, found in *trnQ,* characterized most of Odontoceti (except *P. macrocephalus* + *K. breviceps*) and was followed by successive HCBCs.

The analysis of the distribution of the CSBPSs revealed the occurrence of a dynamic, continuous, and still ongoing evolutionary mechanism of changes on the stems of tRNAs. The oldest CSBPSs were followed by successive changes (marked with an asterisk in Fig 5) that occurred in descendant groups at different taxonomic ranks (S7-S9 Figs). The CSBPSs were in several cases molecular signatures for the different clades as shown in the examples described below a referred for simplicity to a single tRNA.

The A–T vs. G–C, a FCBC of type I occurring at the 1dh-pair of *trnH,* distinguished the Ziphiidae from other Cetacea (Fig 5; S7 Fig). Similarly the [5'(G vs. A), 3' (T)] HCBC, observed in the 7ac-pair of *trnH*, differentiated *P. blainvillei* from other cetacean taxa (Fig 5; S8 Fig, S10 tRNAs multiple alignments, *trnH*). Finally, the T|C vs. T–A, a mismatch located at the 3'end of the 2tp-pair of the same tRNA, sets apart Delphinidae from other Cetacea (Fig 5; S9 Fig).

In other cases, the CSBPSs represented events of convergent/parallel evolution. Thus, the T–A vs. C–G FCBC, located at the tp4-pair of *trnF*, was shared by *Balaena misticetus* and *Eschrichtius robustus* (Fig 5; S7 Fig, S10 tRNAs multiple alignments, *trnF*). Analogously, *Neophocaena phocaenoides, O. brevirostris, Feresa attenuata* and *O. orca* complex exhibited the same HCBC [5'(G vs. A), 3' (T)], at the 4ac-pair of *trnD* (Fig 5; S8 Fig, S10 tRNAs multiple alignments, *trnD*). Finally, the T|T vs. A–T mismatch, at the 6ac-pair of the *trnC*, appeared independently in *Sousa chinensis* and *O. orca* complex (Fig 5; S9 Fig, S10 tRNAs multiple alignments, *trnC*).

The substitution pattern produced in some cases the secondary reversion to the condition observed in outgroups, due to the limited possibility of combinations of the four bases. Thus, in *M. monoceros* a

secondary A–T vs. G–C FCBC occurred in the 5ac-pair of *trnW* (Fig 5; S7 Fig, S10 tRNAs multiple alignments, *trnW*). Similar reversions were observed also for HCBCs and mismatches.

The distribution of CSBPSs in living Cetacea exhibited a broad range of variation (Fig 5; S7-S9 Figs, S10 tRNAs multiple alignments). Most of Mysticeti showed a smaller number CSBPSs than Odontoceti. *E. robustus* (10 CSBPSs) and, more markedly, *C. marginata* (19 CSBPSs) were two exceptions to this behaviour. Within Odontoceti, most of Delphinidae exhibited a lower number of CSBPSs than members of other families. *P. blainvillei* and *L. vexillifer* showed the maximum number (30) of CSBPSs. Other species with at least 20 CSBPSs were *Berardius bairdii* (20), *I. geoffrensis* (23), *K. breviceps* (24), *Pl. minor* (28), and *P. macrocephalus* (26). *O. orca* complex was a peculiar case to be analysed. If this taxon was considered as an assembly of multiple cryptic-species a very low number of CSBPSs could be detected. Conversely, if different specimens of *O. orca* were considered to be derived from a single species, then a minimum of 17 CSBPSs could be assigned to this taxon (Fig 5).

Comparison of the reference topology (S1 Fig) with the distribution of CSBPSs (Fig 5) showed that a good agreement existed among the lengths of the branches and the numbers of CSBPSs.

The distribution of CSBPSs was further investigated in the ten species of Odontoceti with a minimum of 4 FCBCs (Fig 5). None tRNA, exhibiting at least one FCBC, was shared by all taxa. However, when a tRNA was shared by different species, a FCBC could be located in the same stem-pair in different taxa (e.g. 5ac-pair of *trnH* in *P. blainvillei and P. macrocephalus*) (S7 Fig, S10 tRNAs multiple alignments). Similar patterns were detected for HCBCs and mismatches (S8 and S9 Figs, S10 tRNAs multiple alignments).

M. grayi and *N. asiaorientalis* were not considered in most of the analyses performed in present paper (see above Materials and methods). However, it was possible to include the tRNAs of these species in the multiple alignments (S10 tRNAs multiple alignments).

The analysis of these alignments allowed to identify for *M. grayi* at least 5 FCBCs (*trnH*, 3tp-pair; *trnN*,; *trnQ*, 1-an-pair; *trnT*, 1an-pair; *trnW*, 5tp-pair), 3 HCBCs (*trnC*, 5’ end of 1ac-pair; *trnF*, 5’ end of 3an-pair; *trnQ*, 5’end of 3an-pair) and 2 mismatches (*trnL2*, 5’ end of 7ac-pair; *trnV*, 3’ end of 4an-pair). Particularly interesting was the presence in the 2tp-pair of *trnN* of a FCBC of type II (i.e. A–T vs. T–A). The total number of 10 CSBPSs found in *M. grayi* was in agreement with values obtained for other *Mesoplodon* species (Fig 5).

Two HCBCs resulted unquestionably peculiar to *N. asiaorientalis*. They were located respectively at the 5tp-pair of *trnD* [(G), (T vs. C)], and at 2tp-pair of *trnS2* [5’(A vs. G), 3’(T)]. Furthermore, *N. asiaorientalis* shared with *N. phocaenoides* most of CSBPSs recorded for this latter species (Fig 5; S10 tRNAs multiple alignments).

Intraspecific variation of CSBPSs

The intraspecific level of CSBPSs variation was studied in nine cetacean species (S2 Table). The analyses were performed for *Balaenoptera physalus*, *M. densirostris*, *Mesoplodon europaeus*, *O. orca*, *P. macrocephalus*, *Tursiops aduncus*, *Tursiops australis*, *Tursiops truncates,* and *Z. cavirostris*. Intraspecific FCBCs were not identified, irrespective to the number of mtDNAs analysed (8–152). Conversely, a variable, but limited, number of HCBCs and mismatches was detected. *B. physalus* exhibited intraspecific HCBCs and mismatches in 13 tRNAs. *M. densirostris* presented CSBPSs in six tRNAs, while *M. europaeus* showed one HCBC in *trnL2*. In the *O. orca* complex intraspecific CSBPSs, not present in 7 sequences included in 94T-set, were found in seven tRNAs. *P. macrocephalus* exhibited CSBPSs in four tRNAs. *T. aduncus,* showed one HCBC in *trnF* and *trnT*. *T. australis* presented a mismatch in *trnS1*. In *T. truncatus* CSBPSs were found in *trnD*, *trnF*, *trnL2*, *trnR*, and *trnV*. Finally in *Z. cavirostris* a single HCBC was identified in *trnD*, *trnH*, *trnK*, *trnL2*, *trnS1*, and *trnT*. As a general behaviour, the intraspecific variation of CSBPSs resulted limited.

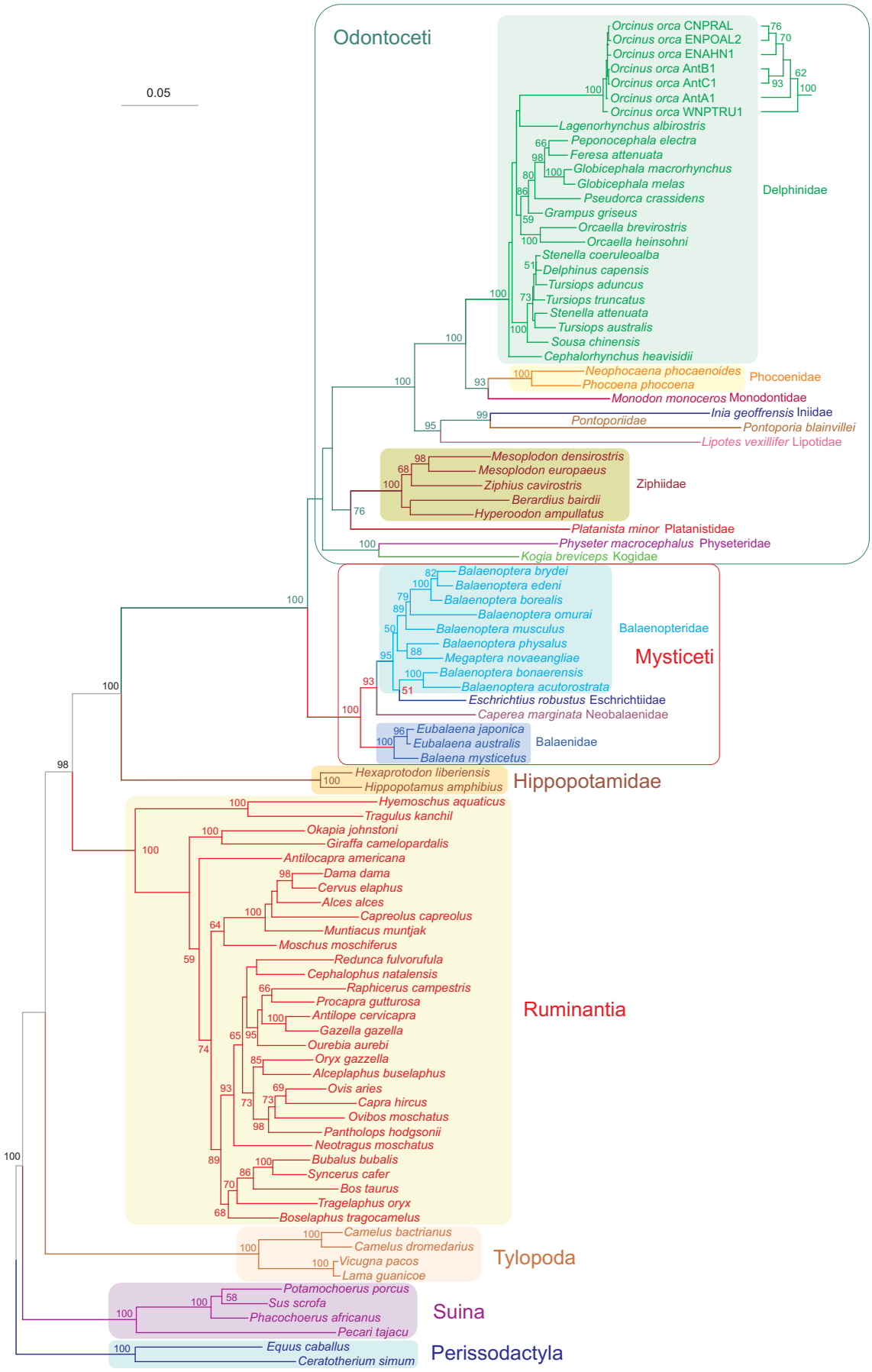


Figure S1. The phylogeny of Cetartiodactyla.
Maximum likelihood (-lnL = 66963.002109) phylogram depicting the phylogenetic relationships among the major clades of Cetartiodactyla. The tree was created by analysing the amino acid 94T-set (3716 positions) with the RAxML 7.4.2 program implemented in raxmlGUI 1.3.1. The evolutionary model was MTMAM + F + CAT. Thirteen partitions were applied: one for every protein. The numbers represent bootstrap values expressed in percent. Only bootstrap values ≥ 50% are provided for the nodes. The scale bar represents 0.05 substitutions/site.

Ziphius cavirostris mtDNA — 16552 bp

	start	end	size		
cox1	1	1551	1551	ATG	AGA
trnS2	1547	1615	69	tga	
isp (*trnS2*-*trnD*)	1616	1622	7		
trnD	1623	1690	68	gtc	
cox2	1691	2374	684	ATG	TAA
isp (*cox2*-*trnK*)	2375	2377	3		
trnK	2378	2444	67	ttt	
isp (*trnK*-*atp8*)	2445	2445	1		
atp8	2446	2646	201	ATG	TAA
atp6	2607	3286	680	ATG	TA(a)
cox3	3287	4071	785	ATG	TA(a)
trnG	4072	4140	69	tcc	
nad3	4141	4487	347	ATA	TA(a)
trnR	4488	4557	70	tcg	
nad4L	4558	4854	297	GTG	TAA
nad4	4848	6225	1378	ATG	T(aa)
trnH	6226	6295	70	gtg	
trnS1	6296	6355	60	gct	
isp (*trnS1*-*trnL1*)	6356	6356	1		
trnL1	6357	6426	70	tag	
nad5	6427	8247	1821	ATA	TAA
nad6	8231	8758	528	ATG	TAA
trnE	8759	8827	69	ttc	
isp (*trnE*-*cob*)	8828	8831	4		
cob	8832	9971	1140	ATG	AGA

	start	end	size		
trnT	9972	10042	71	tgt	
trnP	10042	10108	67	tgg	
Control Region	10109	10989	881		
trnF	10990	11063	74	gaa	
rrnS	11064	12035	972		
trnV	12036	12102	67	tac	
rrnL	12103	13682	1580		
trnL2	13683	13757	75	taa	
isp (*trnL2*-*nad1*)	13758	13759	2		
nad1	13760	14716	957	ATG	TAG
isp (*nad1*-*trnI*)	14717	14720	4		
trnI	14721	14789	69	gat	
trnQ	14787	14859	73	ttg	
isp (*trnQ*-*trnM*)	14860	14860	1		
trnM	14861	14929	69	cat	
nad2	14930	15971	1042	ATG	T(aa)
trnW	15972	16038	67	tca	
isp (*trnW*-*trnA*)	16039	16042	4		
trnA	16043	16111	69	tgc	
isp (*trnA*-*trnN*)	16112	16112	1		
trnN	16113	16186	74	gtt	
OL	16187	16218	32		
trnC	16219	16285	67	gca	
trnY	16286	16351	66	gta	
isp (*trnY*-*cox1*)	16352	16352	1		

Figure S2. The mitochondrial genome of *Ziphius cavirostris*.
Tthe gene order is depicted and linearised starting from ***cox1***. Genes encoded on the **α-strand** (right to left orientation) are underlined in **green**, while those encoded on the **β-strand** are underlined in **red** (left to right orientation). Gene nomenclature: ***atp6*** and ***atp8***: ATP synthase subunits 6 and 8; ***cob***: apocytochrome b; ***cox1-3***: cytochrome c oxidase subunits 1–3; ***nad1-6*** and ***nad4L***: NADH dehydrogenase subunits 1–6 and 4L; ***rrnS*** and ***rrnL***: small and large subunit ribosomal RNA (rRNA) genes; and X: transfer RNA (tRNA) genes, where X is the one-letter abbreviation of the corresponding amino acid. In particular, L1 identifies the CTN codon family, L2 the TTR codon family, S1 the AGY codon family, and S2 the TCN codon family. **CR**, Control Region. **OL**, origin of duplication of the light strand. A black circle located between two adjacent genes denotes the presence of an intergenic spacer (in white is the number of nucleotides forming the spacer). A white circle located between two adjacent genes denotes the presence of an overlapping segment (in red is the number of nucleotides forming the segment). **Isp**, intergenic spacer; **start**, start of the gene; **end**, end of the gene; **size**, size of the gene. For protein-coding genes, the **start codon** is provided in **cyan** and the **stop codon** in **red** (with incomplete stop codons written in parentheses). The anticodon is provided for every tRNA (e.g., **tga** for ***trnS2***).

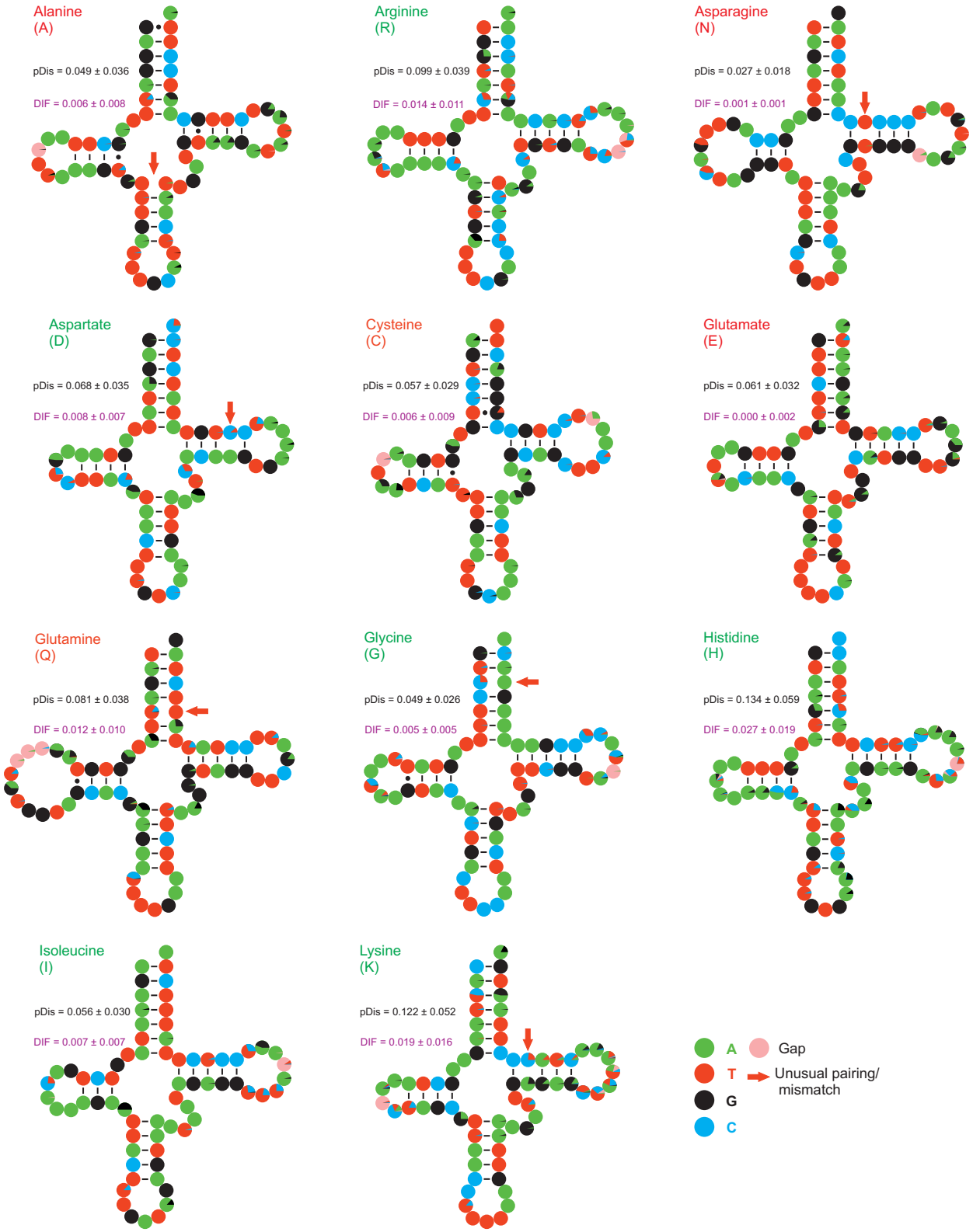


Figure S3. Secondary structure of Cetacea tRNA and level of conservation (*trnA*-*trnK*).
pDis, p-Distance calculated for each pairwise-comparison orthologous tRNAs. **MLdis**, maximum composite likelihood distance calculated for every pairwise-comparison. **DIF** (**MLdis**–**pDis**), the difference between **MLdis** and **pDis**. The **average values** and the **standard deviation** are provided for both **pDis** and **DIF**. The values were computed for each set of orthologous tRNAs.

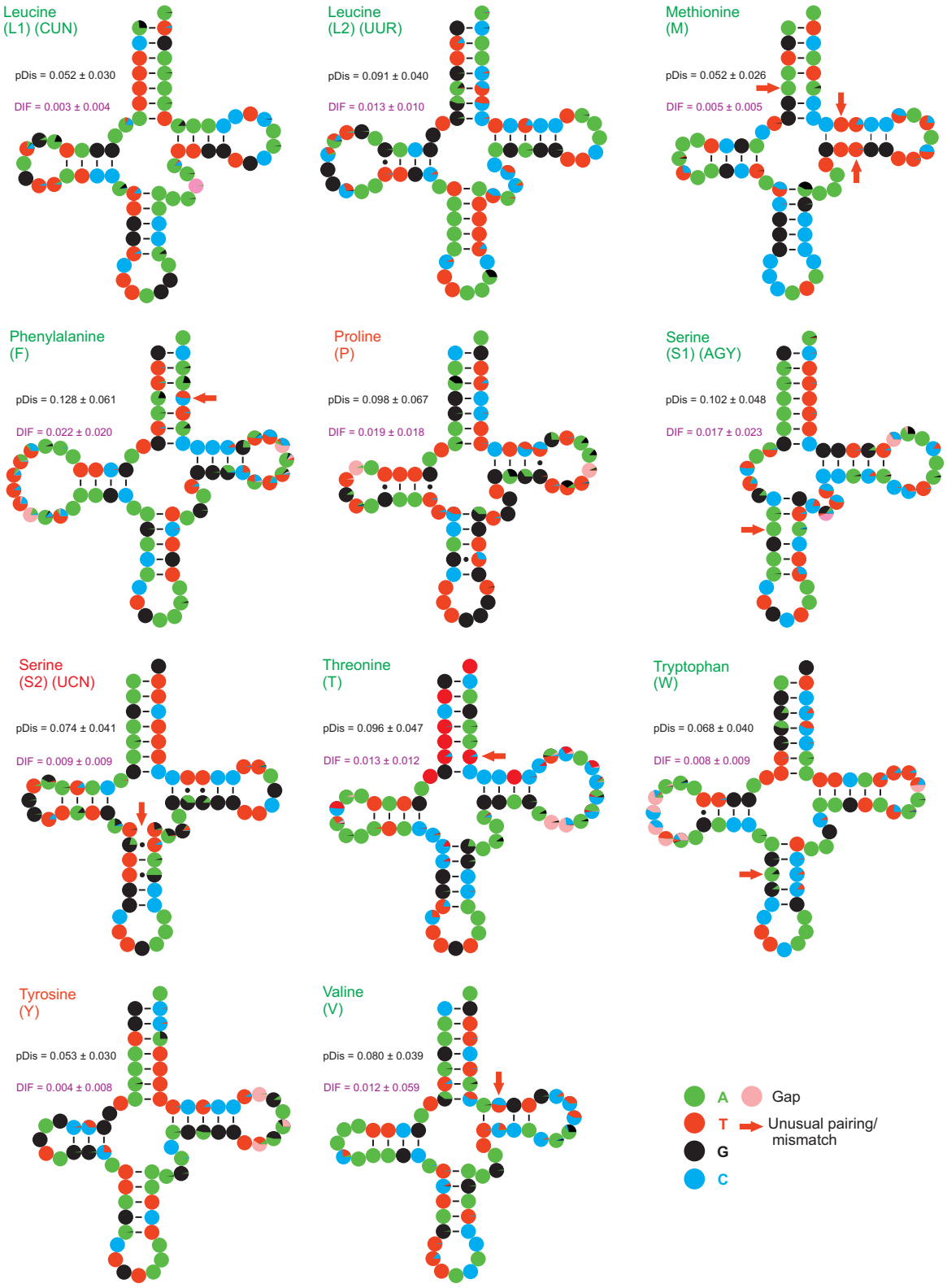


Figure S4. Secondary structure of Cetacea tRNA and level of conservation (*trnL1*-*trnV*).
pDis, p-Distance calculated for each pairwise-comparison orthologous tRNAs. **MLdis**, maximum composite likelihood distance calculated for every pairwise-comparison. **DIF** (**MLdis**–**pDis**), the difference between **MLdis** and **pDis**. The **average values** and the **standard deviation** are provided for both **pDis** and **DIF**. The values were computed for each set of orthologous tRNAs.

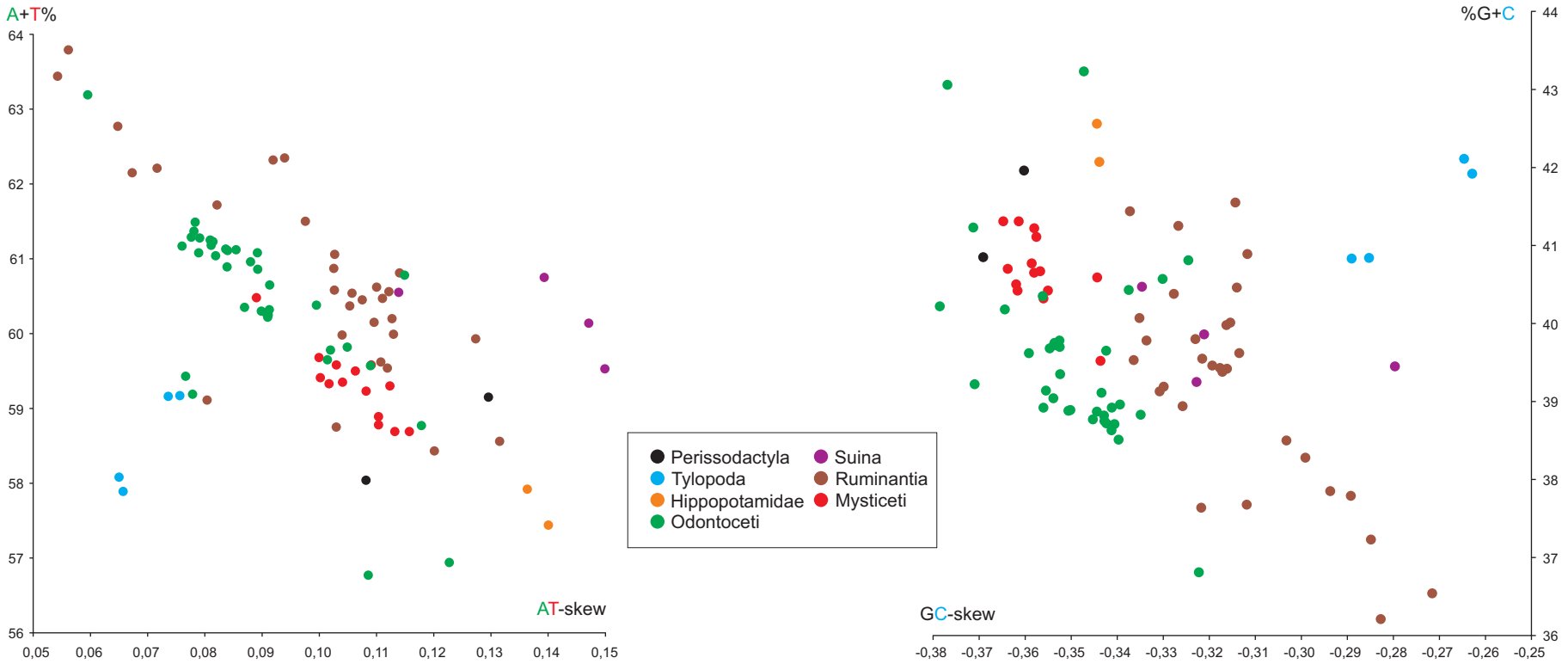


Figure S5. AT-skew vs. A+T% and GC-skew vs. G+C% in the 94T-set mtDNAs.
The values were calculated on the **α-strand** of the full-length mtDNA genomes. The X axis provides the skews values, while the Y axis provides the A+T% and G+C% values.

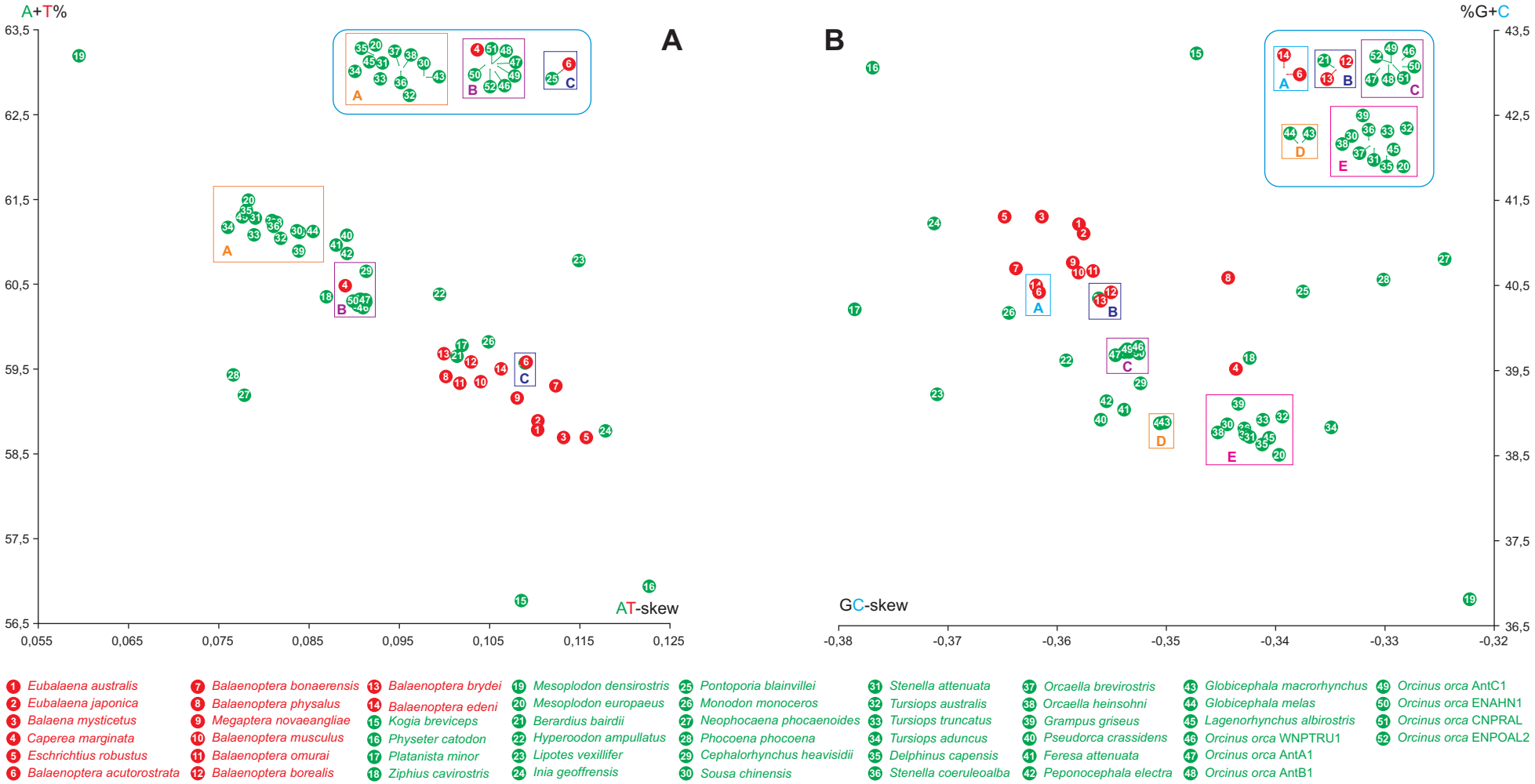


Figure S6. AT-skew vs. A+T% (A), and GC-skew vs. G+C% (B) in Cetacea mtDNAs.
The values were calculated on **α-strand** of the full-length mtDNA genomes. The X axis provides the **AT-** and **GC-**skews values, while the Y axis provides the **A+T%,G+C%** values. Species with a placement that is difficult to identify in the main plots are depicted in the frames.

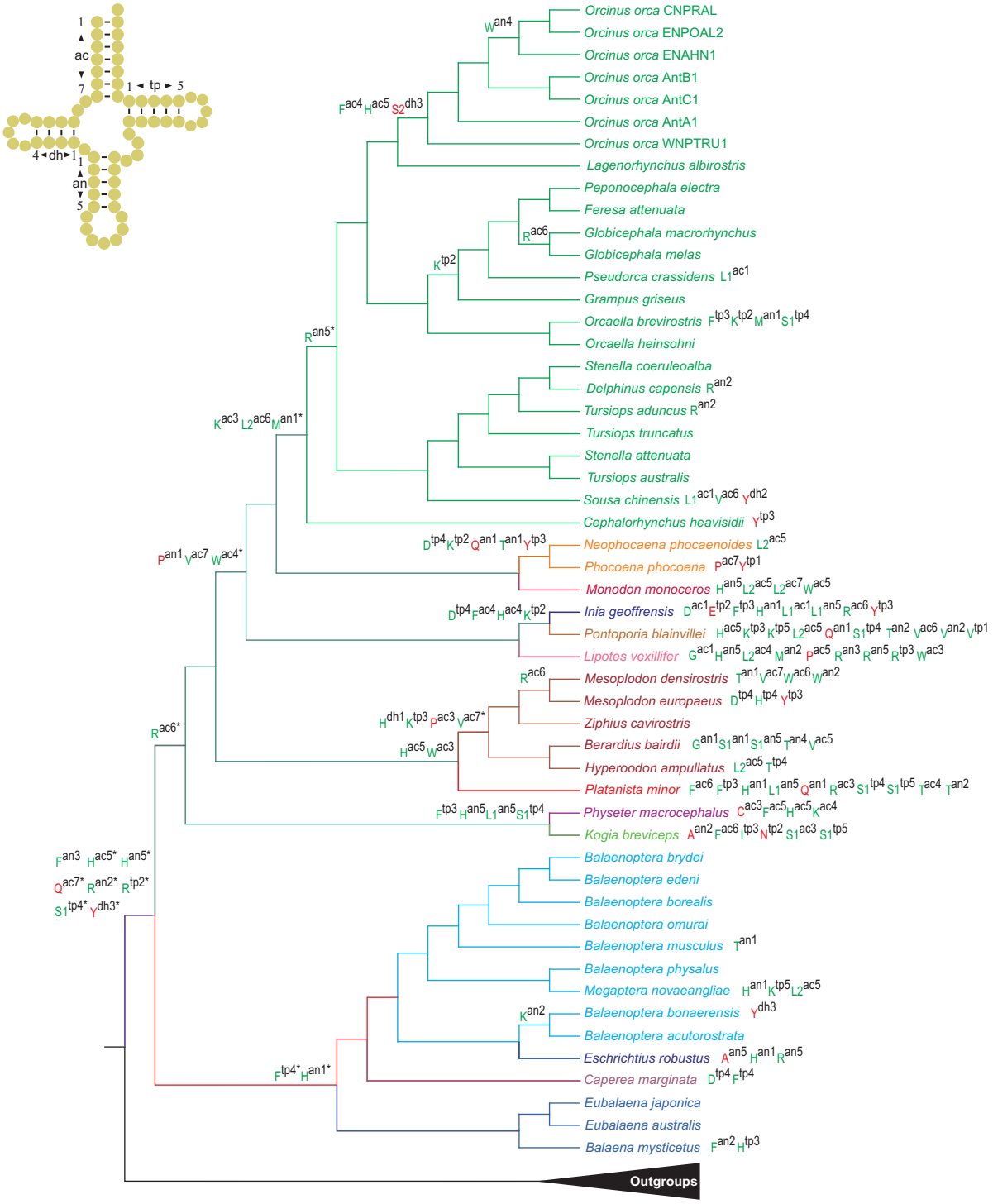


Figure S7. Mapping of FCBCs on the Cetacea phylogenetic tree.
FCBC, fully compensatory base change; **SPIC**, stem position involved in base change. The tRNA and the stem pair involved in **FCBCs** are mapped on the corresponding nodes of the reference phylogenetic tree. The tRNAs are depicted with the single-letter IUPAC code used for the corresponding amino acid. In particular, **L1** identifies the CTN codon family, **L2** the TTR codon family, S1 the AGY codon family, and **S2** the TCN codon family. The stem pair involved in **FCBC** is provided in superscript. The **asterisk**, associated with some **FCBCs** indicates that these **FCBCs** were subjected to successive changes in one/some of the taxa located downstream of the considered node.

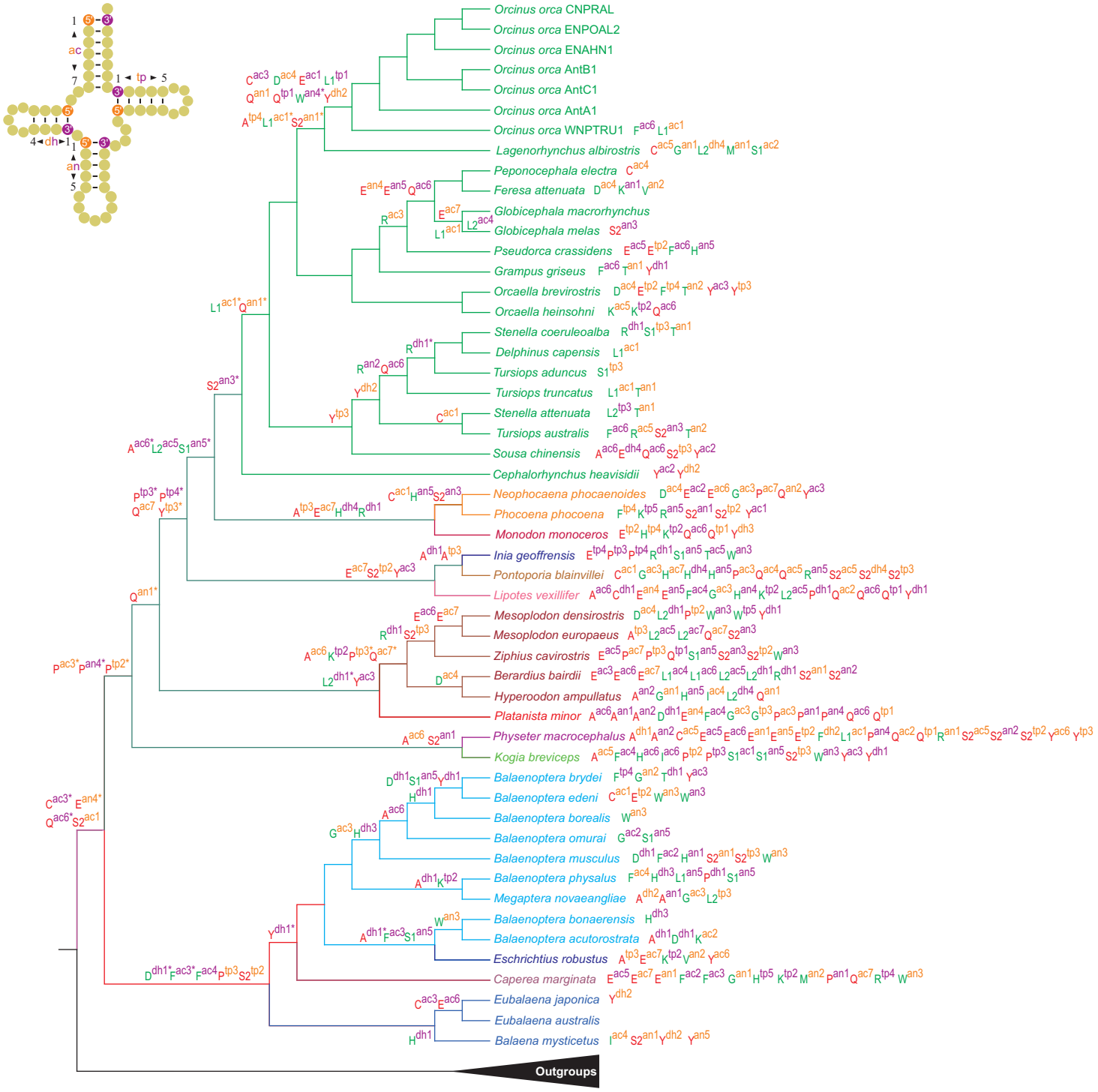


Figure S8. Mapping of HFBCs on the Cetacea phylogenetic tree.
HFBC, hemi-compensatory base change; **SPIC**, stem position involved in base change. The tRNA and the **SPICs** involved in **HFBCs** are mapped on the corresponding nodes of the reference phylogenetic tree. The tRNAs are depicted with the single-letter IUPAC code used for the corresponding amino acid. In particular, **L1** identifies the CTN codon family, **L2** the TTR codon family, **S1** the AGY codon family, and **S2** the TCN codon family. The **SPIC** involved in **HFBC** is provided in **superscript**. The **asterisk** associated with some **HFBCs** indicates that these **HFBCs** were subjected to successive changes in one/some of the taxa located downstream of the considered node. A **SPIC** located on the **5' side** of a stem-pair is marked in **orange**, while a **SPIC** placed on the **3' end** of a pair is **purple**.

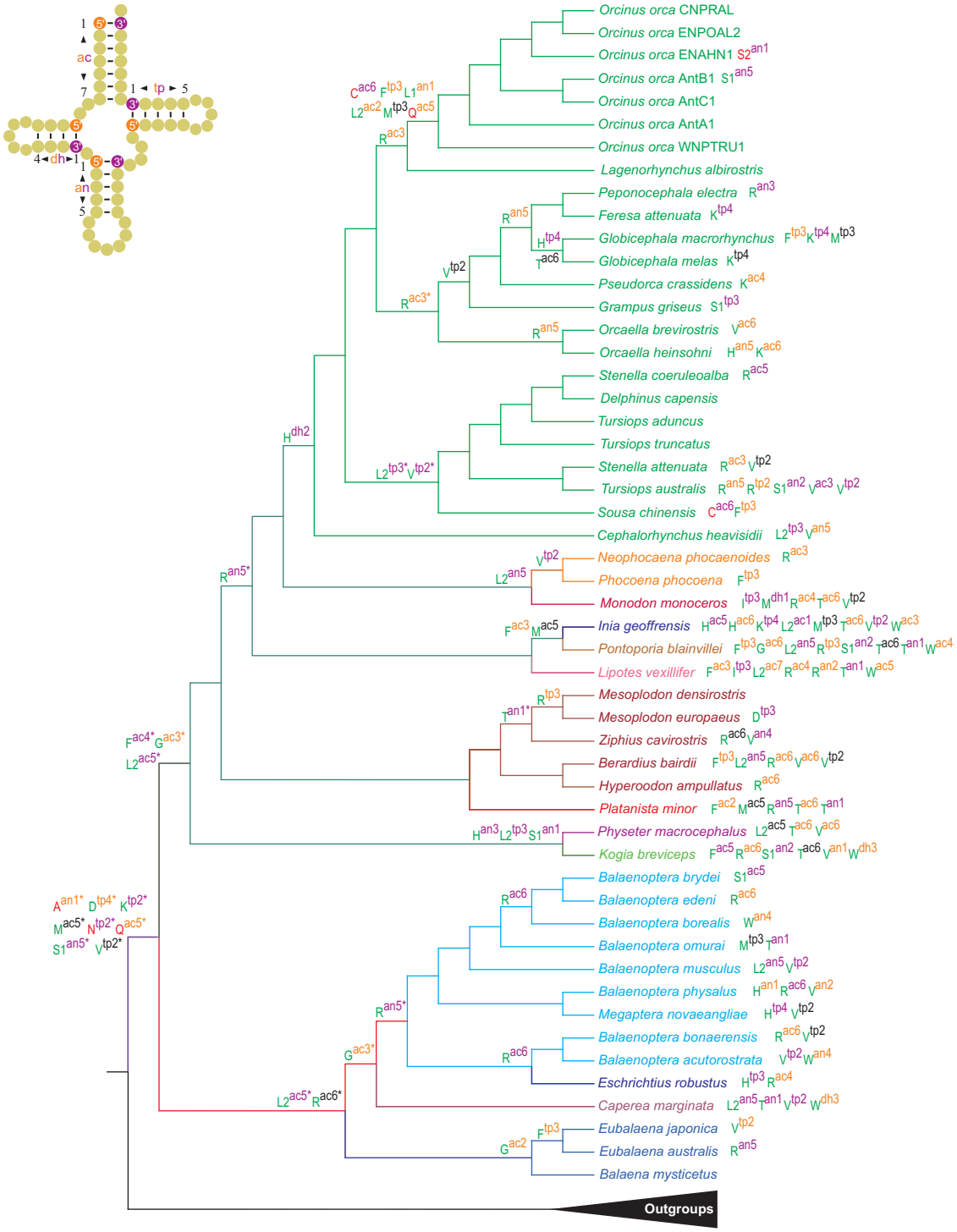


Figure S9. Mapping of mismatches on the Cetacea phylogenetic tree.
Mismatch, **mismatch** in a base pair of a stem; **SPIC**, stem position involved in base change. The tRNAs and **SPICs** involved in **mismatches** are mapped on the corresponding nodes of the reference phylogenetic tree. The tRNAs are depicted with the single-letter IUPAC code used for the corresponding amino acid. In particular, **L1** identifies the CTN codon family, **L2** the TTR codon family, **S1** the AGY codon family, and **S2** the TCN codon family. The **SPIC** involved in a **mismatch** is provided in **superscript**. The **asterisk** associated with some **mismatches** indicates that these **mismatches** were subjected to successive changes in one/some of the taxa located downstream of the considered node. A **SPIC** located on the **5' side** of a stem-pair is marked in **orange**, and a **SPIC** placed on the **3' end** of a pair is **purple**.

Table S1. List of taxa, accession number in GenBank and reference (1/3)

			Perissodactyla		
		Equidae	*Equus caballus* Linnaeus, 1758	X79547	Xu and Arnason 1994
		Rhinocerotidae	*Ceratotherium simum* (Burchell, 1817)	Y07726	Xu and Arnason 1997
			Cetartiodactyla Montgelard, Catzeflis and Douzery, 1997		
Suina		Suidae	*Phacochoerus africanus* (Gmelin, 1788)	DQ409327	Wu et al. 2007
Suina		Suidae	*Potamochoerus porcus* (Linnaeus, 1758)	JN632688	Hassanin et al. 2012
Suina		Suidae	*Sus scrofa* Linnaeus, 1758	FJ237000	Unpublised; Alves and Fernandez
Suina		Tayassuidae	*Pecari tajacu* (Linnaeus, 1758)	AP003427	Unpublished; Yasue et al.
Tylopoda		Camelidae	*Camelus bactrianus* Linnaeus, 1758	EF212037	Ji et al. 2009
Tylopoda		Camelidae	*Camelus dromedarius* Linnaeus, 1758	JN632608	Hassanin et al. 2012
Tylopoda		Camelidae	*Lama guanicoe* (Müller, 1776)	EU681954	Di rocco et al. 2010
Tylopoda		Camelidae	*Vicugna pacos* (Linnaeus, 1758)	NC_002504	Ursing et al. 2000
Ruminantia	Pecora	Antilocapridae	*Antilocapra americana* Ord, 1815	JN632597	Hassanin et al. 2012
Ruminantia	Pecora	Bovidae Alcelaphinae	*Alceplaphus buselaphus* (Pallas, 1766)	JN632593	Hassanin et al. 2012
Ruminantia	Pecora	Bovidae Antilopinae	*Antilope cervicapra* (Linnaeus, 1758)	JN632598	Hassanin et al. 2012
Ruminantia	Pecora	Bovidae Antilopinae	*Gazella gazella* (Pallas, 1766)	JN632640	Hassanin et al. 2012
Ruminantia	Pecora	Bovidae Antilopinae	*Neotragus moschatus* Von Dueben, 1846	JN632669	Hassanin et al. 2012
Ruminantia	Pecora	Bovidae Antilopinae	*Ourebia aurebi* (Zimmermann, 1783)	JN632680	Hassanin et al. 2012
Ruminantia	Pecora	Bovidae Antilopinae	*Pantholops hodgsonii* (Abel, 1826)	DQ191826	Xu et al. 2005
Ruminantia	Pecora	Bovidae Antilopinae	*Procapra gutturosa* (Pallas, 1777)	JN632689	Hassanin et al. 2012
Ruminantia	Pecora	Bovidae Antilopinae	*Raphicerus campestris* (Thunberg, 1811)	JN632693	Hassanin et al. 2012
Ruminantia	Pecora	Bovidae Bovinae	*Bos taurus* Linnaeus, 1758	AY526085	Unpublished; Chung and Ha
Ruminantia	Pecora	Bovidae Bovinae	*Boselaphus tragocamelus* (Pallas, 1766)	EF536350	Hassanin et al. 2012
Ruminantia	Pecora	Bovidae Bovinae	*Bubalus bubalis* (Linnaeus, 1758)	AF547270	Unpublished; Verma et al
Ruminantia	Pecora	Bovidae Bovinae	*Syncerus cafer* (Sparrman, 1779)	EF536353	Hassanin et al. 2012
Ruminantia	Pecora	Bovidae Bovinae	*Tragelaphus oryx* (Pallas, 1766)	JN632704	Hassanin et al. 2012
Ruminantia	Pecora	Bovidae Caprinae	*Capra hircus* Linnaeus, 1758	GU295658	Hassanin et al. 2010
Ruminantia	Pecora	Bovidae Caprinae	*Ovibos moschatus* (Zimmermann, 1780)	FJ207536	Hassanin et al. 2009
Ruminantia	Pecora	Bovidae Caprinae	*Ovis aries* Linnaeus, 1758	AF010406	Hiendleder et al. 1998
Ruminantia	Pecora	Bovidae Cephalophinae	*Cephalophus natalensis* A. Smith, 1834	JN632618	Hassanin et al. 2012
Ruminantia	Pecora	Bovidae Hippotraginae	*Oryx gazzella* (Linnaeus, 1758)	JN632678	Hassanin et al. 2012
Ruminantia	Pecora	Bovidae Reduncinae	*Redunca fulvorufula* (Afzelius, 1815)	JN632695	Hassanin et al. 2012
Ruminantia	Pecora	Cervidae Capreolinae	*Alces alces* (Linnaeus, 1758)	JN632595	Hassanin et al. 2012
Ruminantia	Pecora	Cervidae Capreolinae	*Capreolus capreolus* (Linnaeus, 1758)	JN632610	Hassanin et al. 2012
Ruminantia	Pecora	Cervidae Cervinae	*Cervus elaphus* Linnaeus, 1758	AB245427	Unpublished; Wada et al
Ruminantia	Pecora	Cervidae Cervinae	*Dama dama dama* (Linnaeus, 1758)	JN632629	Hassanin et al. 2012
Ruminantia	Pecora	Cervidae Muntiacinae	*Muntiacus muntjak* (Zimmermann, 1780)	AY225986	Unpublished; Shi et al
Ruminantia	Pecora	Giraffidae	*Giraffa camelopardalis* (Linnaeus, 1758)	JN632645	Hassanin et al. 2012
Ruminantia	Pecora	Giraffidae	*Okapia johnstoni (*P. L. Sclater, 1901)	JN632674	Hassanin et al. 2012
Ruminantia	Pecora	Moschidae	*Moschus moschiferus* Linnaeus, 1758	JN632662	Hassanin et al. 2012
Ruminantia	Tragulina	Tragulidae	*Hyemoschus aquaticus* (Ogilby, 1841)	JN632650	Hassanin et al. 2012
Ruminantia	Tragulina	Tragulidae	*Tragulus kanchil* Raffles, 1821	JN632709	Hassanin et al. 2012
		Hippopotamidae	*Hexaprotodon liberiensis* (Morton, 1849)	JN632625	Hassanin et al. 2012
		Hippopotamidae	*Hippopotamus amphibius* Linnaeus, 1758	AJ010957	Ursing and Arnason 1998

Table S1. List of taxa, accession number in GenBank and reference (2/3)

Cetacea	Mysticeti	Balaenidae	*Balaena mysticetus* Linnaeus, 1758	AJ554051	Arnason et al. 2004
Cetacea	Mysticeti	Balaenidae	*Eubalaena australis* (Gray, 1821)	AP006473	Sasaki et al. 2005
Cetacea	Mysticeti	Balaenidae	*Eubalaena japonica* (Lacépède, 1818)	AP006474	Sasaki et al. 2005
Cetacea	Mysticeti	Balaenopteridae	*Balaenoptera acutorostrata* Lacépède, 1804	AJ554054	Arnason et al. 2004
Cetacea	Mysticeti	Balaenopteridae	*Balaenoptera bonaerensis* Burmeister, 1867	AP006466	Sasaki et al. 2005
Cetacea	Mysticeti	Balaenopteridae	*Balaenoptera borealis* Lesson, 1828	AP006470	Sasaki et al. 2005
Cetacea	Mysticeti	Balaenopteridae	*Balaenoptera brydei* Olsen, 1913	AP006469	Sasaki et al. 2005
Cetacea	Mysticeti	Balaenopteridae	*Balaenoptera edeni* Anderson, 1879	AB201258	Sasaki et al. 2006
Cetacea	Mysticeti	Balaenopteridae	*Balaenoptera musculus* (Linnaeus, 1758)	X72204	Arnason and Gullberg 1993
Cetacea	Mysticeti	Balaenopteridae	*Balaenoptera omurai* Wada et al., 2003	AB201256	Sasaki et al. 2006
Cetacea	Mysticeti	Balaenopteridae	*Balaenoptera physalus* (Linnaeus, 1758)	X61145	Valverde et al. 1994
Cetacea	Mysticeti	Balaenopteridae	*Megaptera novaeangliae* (Borowski, 1781)	AP006467	Sasaki et al. 2005
Cetacea	Mysticeti	Eschrichtiidae	*Eschrichtius robustus* (Lilljeborg, 1860)	AJ554053	Arnason et al. 2004
Cetacea	Mysticeti	Neobalaenidae	*Caperea marginata* (Gray, 1846)	AJ554052	Arnason et al. 2004
Cetacea	Odontoceti	Delphinidae	*Cephalorhynchus heavisidii* (Gray, 1828)	JN632624	Hassanin et al. 2012
Cetacea	Odontoceti	Delphinidae	*Delphinus capensis* Gray, 1828	EU557094	Xiong et al. 2009
Cetacea	Odontoceti	Delphinidae	*Feresa attenuata* Gray, 1874	JF289171	Vilstrup et al. 2011
Cetacea	Odontoceti	Delphinidae	*Globicephala macrorhynchus* Gray, 1846	JF339976	Vilstrup et al. 2011
Cetacea	Odontoceti	Delphinidae	*Globicephala melas* (Traill, 1809)	JF339972	Vilstrup et al. 2011
Cetacea	Odontoceti	Delphinidae	*Grampus griseus* (G. Cuvier, 1812)	EU557095	Xiong et al. 2009
Cetacea	Odontoceti	Delphinidae	*Lagenorhynchus albirostris* (Gray, 1846)	AJ554061	Arnason et al. 2004
Cetacea	Odontoceti	Delphinidae	*Orcaella brevirostris* (Owen in Gray, 1866)	JF289177	Vilstrup et al. 2011
Cetacea	Odontoceti	Delphinidae	*Orcaella heinsohni* Beasley, Robertson and Arnold, 2005	JF339977	Vilstrup et al. 2011
Cetacea	Odontoceti	Delphinidae	**a***Orcinus orca* (Ant_A1) Linnaeus, 1758	GU187217	Morin et al. 2010
Cetacea	Odontoceti	Delphinidae	**b***Orcinus orca* (Ant_B1) Linnaeus, 1758	GU187215	Morin et al. 2010
Cetacea	Odontoceti	Delphinidae	**c***Orcinus orca* (Ant_C1) Linnaeus, 1758	GU187210	Morin et al. 2010
Cetacea	Odontoceti	Delphinidae	**d***Orcinus orca* (CNPNRAL) Linnaeus, 1758	GU187189	Morin et al. 2010
Cetacea	Odontoceti	Delphinidae	**e***Orcinus orca* (ENAHN1) Linnaeus, 1758	GU187178	Morin et al. 2010
Cetacea	Odontoceti	Delphinidae	**f***Orcinus orca* (ENPOAL2) Linnaeus, 1758	GU187201	Morin et al. 2010
Cetacea	Odontoceti	Delphinidae	**g***Orcinus orca* (WNPTRU1) Linnaeus, 1758	GU187159	Morin et al. 2010
Cetacea	Odontoceti	Delphinidae	*Peponocephala electra* (Gray, 1846)	JF289175	Vilstrup et al. 2011
Cetacea	Odontoceti	Delphinidae	*Pseudorca crassidens* (Owen, 1846)	JF289173	Vilstrup et al. 2011
Cetacea	Odontoceti	Delphinidae	*Sousa chinensis* (Osbeck, 1765)	EU557091	Xiong et al. 2009
Cetacea	Odontoceti	Delphinidae	*Stenella attenuata* (Gray, 1846)	EU557096	Xiong et al. 2009
Cetacea	Odontoceti	Delphinidae	*Stenella coeruleoalba* (Meyen, 1833)	EU557097	Xiong et al. 2009
Cetacea	Odontoceti	Delphinidae	*Tursiops aduncus* (Ehrenberg, 1833)	EU557092	Xiong et al. 2009
Cetacea	Odontoceti	Delphinidae	*Tursiops australis* Charlton-Robb et al., 2011	KF570364	Moura et al. 2013
Cetacea	Odontoceti	Delphinidae	*Tursiops truncatus* (Montagu, 1821)	EU557093	Xiong et al. 2009
Cetacea	Odontoceti	Iniidae	*Inia geoffrensis* (Blainville, 1817)	AJ554059	Arnason et al. 2004
Cetacea	Odontoceti	Lipotidae	*Lipotes vexillifer* Miller, 1918	AY789529	Yan et al. 2005
Cetacea	Odontoceti	Monodontidae	*Monodon monoceros* Linnaeus, 1758	AJ554062	Arnason et al. 2004

a-g, sequences representative of the main clades identified within the *Orcinus orca* complex (Morin et al. 2010)

Table S1. List of taxa, accession number in GenBank and reference (3/3)

Cetacea	Odontoceti	Phocoenidae	******_Neophocaena asiaeorientalis_ Pilleri & Gihr, 1972	KP170488	Unpublished; Liu et al.
Cetacea	Odontoceti	Phocoenidae	_Neophocaena phocaenoides_ (G. Cuvier, 1829)	KC777291	Unpublished; Xu et al
Cetacea	Odontoceti	Phocoenidae	_Phocoena phocoena_ (Linnaeus, 1758)	AJ554063	Arnason et al. 2004
Cetacea	Odontoceti	Kogidae	_Kogia breviceps_ (Blainville, 1838)	AJ554055	Arnason et al. 2004
Cetacea	Odontoceti	Physeteridae	_Physeter macrocephalus_ Linnaeus, 1758	AJ277029	Arnason et al. 2000
Cetacea	Odontoceti	Platanistidae	_Platanista minor_ Owen, 1853	AJ554058	Arnason et al. 2004
Cetacea	Odontoceti	Pontoporidae	_Pontoporia blainvillei_ (Gervais and d'Orbigny, 1844)	AJ554060	Arnason et al. 2004
Cetacea	Odontoceti	Ziphiidae	_Berardius bairdii_ Stejneger, 1883	AJ554057	Arnason et al. 2004
Cetacea	Odontoceti	Ziphiidae	_Hyperoodon ampullatus_ (Forster, 1770)	AJ554056	Arnason et al. 2004
Cetacea	Odontoceti	Ziphiidae	_Mesoplodon densirostris_ (Blainville, 1817)	KF032861	Unpublished; Morin et al
Cetacea	Odontoceti	Ziphiidae	_Mesoplodon europaeus_ (Gervais, 1855)	KC776688	Unpublished; Morin et al
Cetacea	Odontoceti	Ziphiidae	******_Mesoplodon grayi_ von Haast, 1876	KF981442	Thompson et al 2015.
Cetacea	Odontoceti	Ziphiidae	******* _Mesoplodon ginkgodens_ Nishiwaki and Kamiya, 1958	KR534596	Yao et al 2015
Cetacea	Odontoceti	Ziphiidae	_Ziphius cavirostris_ G. Cuvier, 1823	**LN997430**	**this paper**

******, species used only in selected comparisons and not fully implemented in the analyses; *******, this species become available too late to be implemented in the analyses.

References

Arnason U, Gullberg, A. 1993 Comparison between the complete mtDNA sequences of the blue and the fin whale, two species that can hybridize in nature**.** *J. Mol. Evol.* **37**, 312-322.

Arnason U, Gullberg A, Gretarsdottir S, Ursing B, Janke A. 2000 The mitochondrial genome of the sperm whale and a new molecular reference for estimating eutherian divergence dates. *J. Mol. Evol.* **50**, 569-578.

Arnason U., Gullberg A, Janke A. 2004 Mitogenomic analyses provide new insights into cetacean origin and evolution. *Gene* **333**, 27-34.

Di Rocco F, Zambelli A, Mate L, and Vidal-Rioja L. 2010 The complete mitochondrial DNA sequence of the guanaco (*Lama guanicoe*): comparative analysis with the vicuna (*Vicugna vicugna*) genome**.** *Genetica* **138**, 813-818.

Hassanin A, Ropiquet A, Couloux A, Cruaud C**.** 2009 Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). *J. Mol. Evol.* **68**, 293-310.

Hassanin A, Bonillo C, Nguyen BX, and Cruaud, C**.** 2010 Comparisons between mitochondrial genomes of domestic goat (*Capra hircus*) reveal the presence of numts and multiple sequencing errors**.** *Mitochondrial DNA* **21**, 68-76.

Hassanin A, Delsuc F, Ropiquet A, Hammer C, Jansen van Vuuren B, Matthee C, Ruiz-Garcia M, Catzeflis F, Areskoug V, Nguyen TT, Couloux, A. 2012 Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *C. R. Biol.* **335**, 32-50.

Hiendleder S, Lewalski H, Wassmuth R, Janke A. 1998 The complete mitochondrial DNA sequence of the domestic sheep (*Ovis aries*) and comparison with the other major ovine haplotype. *J. Mol. Evol.* **47**, 441-448.

Ji R, Cui P, Ding F, Geng J, Gao H, Zhang H, Yu J, Hu S, Meng H. 2009 Monophyletic origin of domestic bactrian camel (*Camelus bactrianus*) and its evolutionary relationship with the extant wild camel (*Camelus bactrianus ferus*). *Anim. Genet.* **40**, 377-382.

Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, Durban J, Parsons K, Pitman R, Li L, Bouffard P, Abel Nielsen SC, Rasmussen M, Willerslev E, Gilbert, MT, Harkins T. 2010 Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res.* **20**, 908-916.

Moura AE, Nielsen SC, Vilstrup JT, Moreno-Mayar JV, Gilbert MT, Gray HW, Natoli A, Moller L, Hoelzel AR. 2013 Recent diversification of a marine genus (*Tursiops* spp.) tracks habitat preference and environmental change. *Syst. Biol.* **62**, 865-877.

Sasaki T, Nikaido M, Hamilton H, Goto M, Kato H, Kanda N, Pastene LA, Cao Y, Fordyce RE, Hasegawa M, Okada N. 2005 Mitochondrial phylogenetics and evolution of mysticete whales. *Syst. Biol.* **54**, 77-90.

Sasaki T, Nikaido M, Wada S, Yamada, TK, Cao Y, Hasegawa M, Okada, N. 2006 *Balaenoptera omurai* is a newly discovered baleen whale that represents an ancient evolutionary lineage. *Mol. Phylogenet. Evol.* **41**, 40-52.

Thompson KF, Patel S, Williams L, Tsai P, Constantine R, Baker CS, Millar CD. 2015 High coverage of the complete mitochondrial genome of the rare Gray's beaked whale (*Mesoplodon grayi*) using Illumina next generation sequencing. Mitochondrial DNA. In press; doi:10.3109/19401736.2013.878908

Ursing BM, Arnason U. 1998 Analyses of mitochondrial genomes strongly support a hippopotamus-whale clade. *Proc. R. Soc. Lond., B, Biol. Sci.* **265**, 2251-2255.

Ursing BM, Slack KE, Arnason U. 2000 Subordinal artiodactyl relationships in the light of phylogenetic analysis of 12 mitochondrial protein-coding genes. *Zool. Scr.* **29**, 83-88.

Valverde JR, Marco R, Garesse, R. 1994 A conserved heptamer motif for ribosomal RNA transcription termination in animal mitochondria. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5368-5371.

Vilstrup JT, Ho SY, Foote AD, Morin PA, Kreb D, Krutzen M, Parra GJ, Robertson KM, de Stephanis R, Verborgh P, Willerslev E, Orlando L, Gilbert MT. 2011 Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. *BMC Evol. Biol.* **11**, 65.

Wu GS, Yao YG, Qu KX, Ding ZL, Li H, Palanichamy MG, Duan, ZY, Li N, Chen YS, Zhang YP. 2007 Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in East Asia. *Genome Biol.* **8**, R245.

Xiong Y, Brandley MC, Xu S, Zhou K, Yang G. 2009 Seven new dolphin mitochondrial genomes and a time-calibrated phylogeny of whales. *BMC Evol. Biol.* **9**, 20.

Xu SQ, Yang YZ, Zhou J, Jing GE, Chen YT, Wang J, Yang HM, Wang J, Yu J, Zheng XG, Ge RL. 2005 A mitochondrial genome sequence of the Tibetan antelope (*Pantholops hodgsonii*). *Genomics Proteomics Bioinformatics* **3**, 5-17.

Xu X, Arnason U. 1994 The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region. *Gene* **148**, 357-362.

Xu X, Arnason, U. 1997 The complete mitochondrial DNA sequence of the white rhinoceros, *Ceratotherium simum*, and comparison with the mtDNA sequence of the Indian rhinoceros, *Rhinoceros unicornis*. *Mol. Phylogenet. Evol.* **7**, 189-194.

Yan J, Zhou K, Yang G. 2005 Molecular phylogenetics of 'river dolphins' and the baiji mitochondrial genome. *Mol. Phylogenet. Evol.* **37**, 743-750.

Yao CJ, Chen CH, Hsiao CD. 2015The complete mitogenome of Ginkgo-toothed beaked whale (*Mesoplodon ginkgodens*) (Chordata: Ziphiidae). *Mitochondrial DNA*, 1-2. In press; doi:10.3109/19401736.2015.1053122.

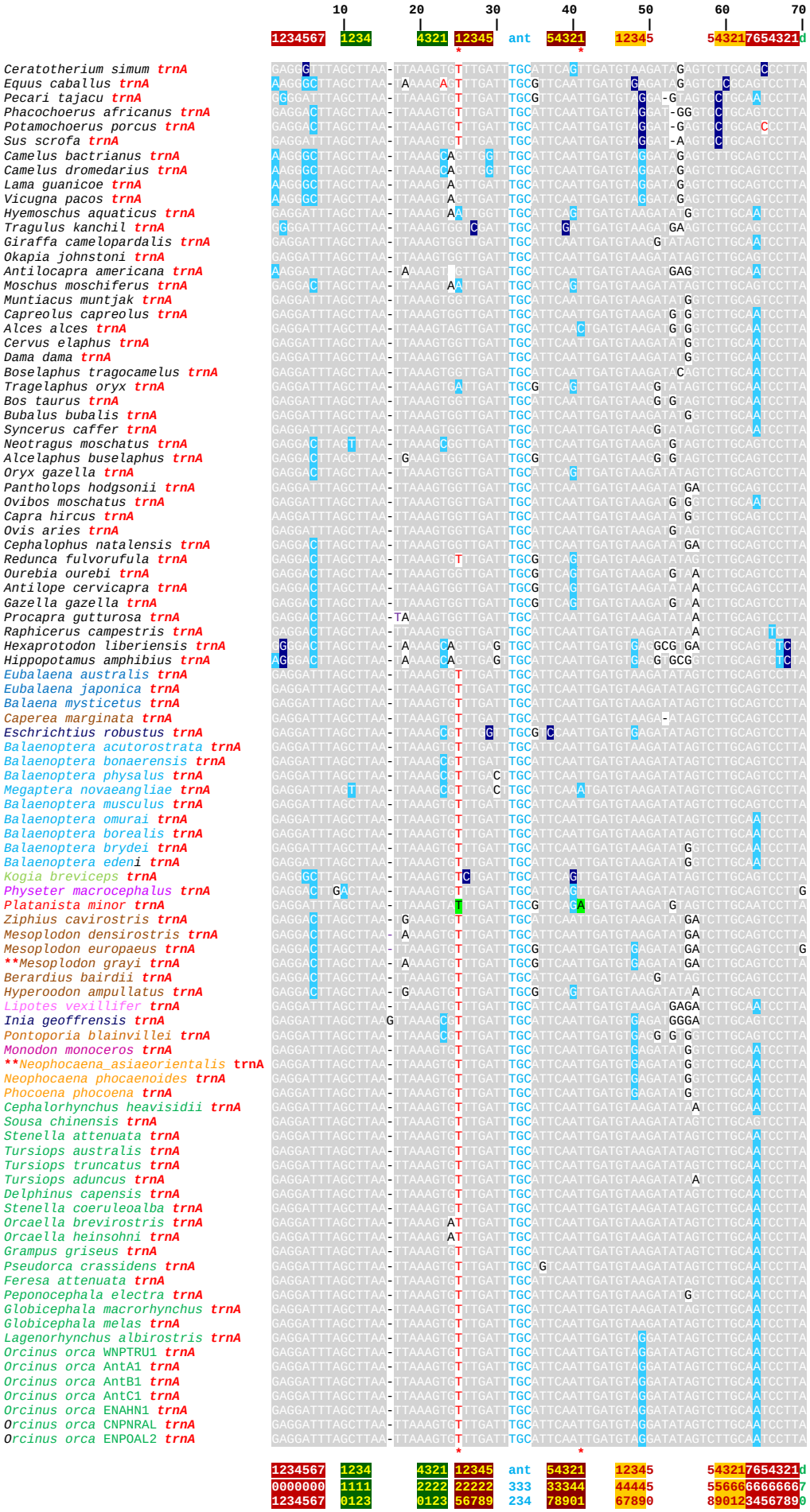
Table S2. Intraspecific CSBPSs identified in some Cetacea

Balaenoptera physalus (152 mtDNAs)	Thirteen tRNAs exhibited intraspecific HCBCs and mismatches in *B. physalus*. Only mismatches were identified in *trnD* (1 seq, 1 mismatch), *trnL2* (5 seqs, 1 mismatch), *trnR* (2 seq, 1 mismatch), *trnS1* (1seq, 1 mismatch), and *trnT* (3 seqs, 3 mismatches). A combination of HCBCs and mismatches was found in *trnF* (4 seqs, 2 HCBC and 2 mismatches) *trnH* (20 seqs with 4 mismatches, 12 seqs with 1 HCBC), *trnP* (2 seqs, 1 HCBC + 1 mismatch). Only HCBCs were identified in *trnA* (4 seqs, 3 HCBCs), *trnC* (2 seqs, 1 HCBC), *trnE* (1 seq, 1 HCBC) *trnL1* (14 seqs, 1 HCBC) *trnN* (1 seq, 1 HCBC). The number of sequences interested by a single CSBPS ranged from 1 to 12.
Mesoplodon densirostris (11 mtDNAs)	Six tRNAs of *M. densirostris* presented CSBPSs i.e. *trnL2* (7 seqs, 1HCBC), *trnQ* (1 seq, 1 HCBC), *trnR* (7 seqs, 1 HCBC), trnS2 (3 seqs, 2 HCBCs), trnV (3 seqs, 2 mismatches), and trnW (7 seqs, 2HCBCs). In *M. europaeus* only trnL2 showed a CSBP (1 seq, 1 HCBC).
Mesoplodon europaeus (11 mtDNAs)	In *M. europaeus* only *trnL2* showed a CSBP (1 seq, 1 HCBC).
Orcinus orca complex (86 mtDNAs)	In the *O. orca* complex we report here only the intraspecific CSBPSs not present in the seven sequences included in 94T-set. Seven tRNAs shown extra CSBPSs i.e. *trnC* (1 seq, 1 HCBC), *trnD* (2 seqs, 1 HCBC), *trnG* (1 seq, 1 mismatch), *trnK* (2 seqs, 1 HCBC), *trnS1* (2 seqs with 1 mismatch, 1 seq with 1 HCBC), *trnT* (1 seq, 1 HCBC), *trnW* (34 seqs, 1 HCBC).
Physeter microcephalus (18 mtDNAs)	Four tRNAs of *P. macrocephalus* exhibited extra CSBPSs i.e. *trnA* (1 seq, 1 HCBC), *trnF* (1 seq, 1 HCBC), *trnS2* (7 seq, 1 HCBC), and *trnT* (1 seq, 1 mismatch).
Tursiops aduncus (20 mtDNAs)	In *T. aduncus*, one HCBC was found in *trnF* (5 seqs) and *trnT* (10 seqs).
Tursiops australis (8 mtDNAs)	In *T. australis* one mismatch was identified in *trnS1* (1 seq).
Tursiops truncatus (48 mtDNAs)	In *T. truncatus* CSBPSs were found in *trnD* (30 seqs, 1 HCBC), *trnF* (2 seqs, 1 mismatch), *trnL2* (1 seq, 1 MISM), *trnR* (2 seqs, 2 mismatches), and *trnV* (1 seq, 1 HCBC).
Ziphius cavirostris (20 mtDNAs)	In *Z. cavirostris* a single HCBC was identified in *trnD* (8 seqs), *trnH* (2 seqs), *trnK* (2 seqs) *trnL2* (3 seqs), *trnS1* (1 seq), and *trnT* (1 seq).

The intraspecific CSBPSs are described following this scheme: tRNA involved (number of sequences (seqs) showing the CSBPS, number and type of CSBPSs associated).

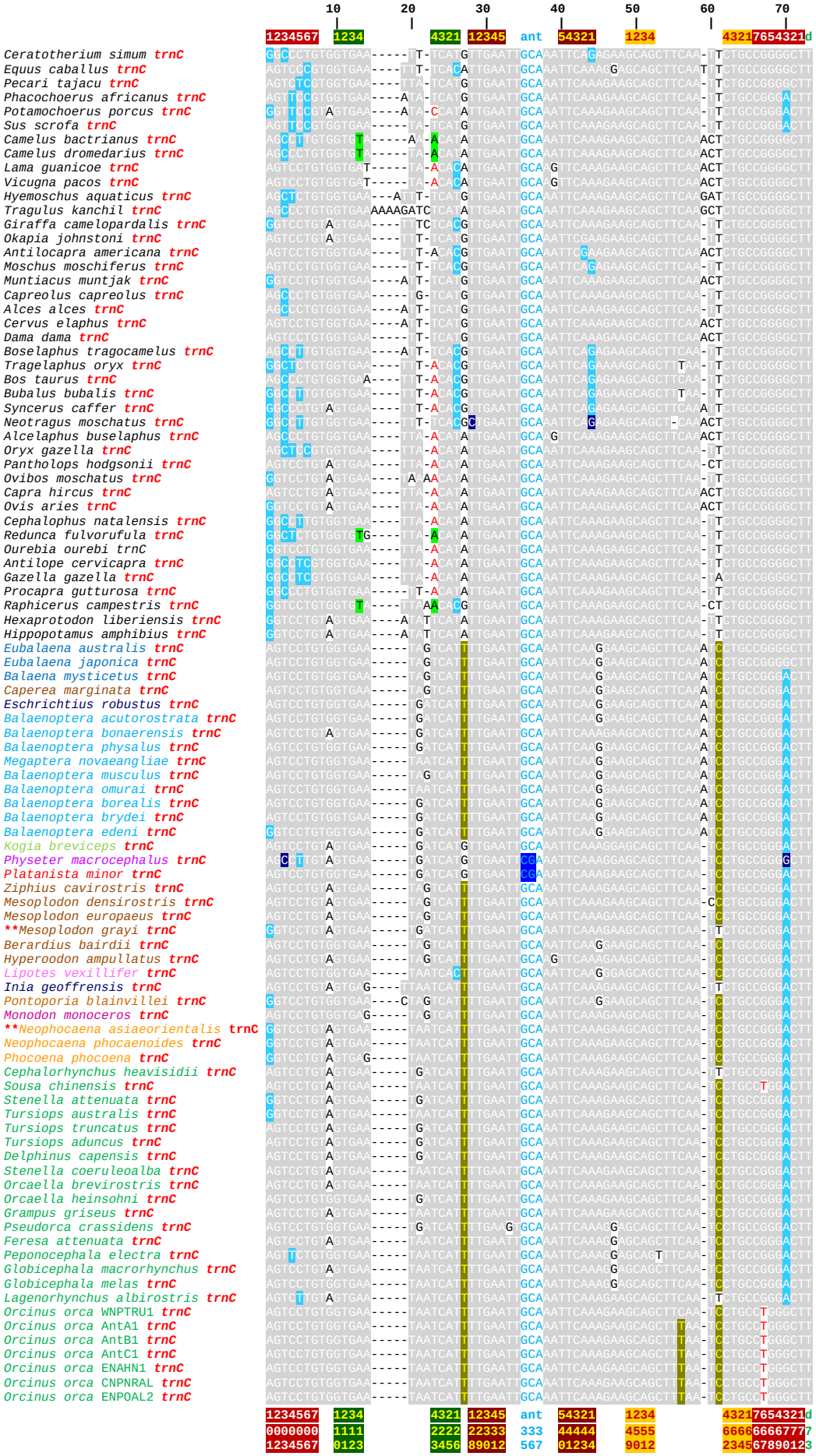
CSBPS, change of sequence in a base pair of a stem; HCBC, hemi-compensatory base change in a base pair of a stem; mismatch, a mismatch in a base pair of a stem.

trnA (ALA) multiple alignment



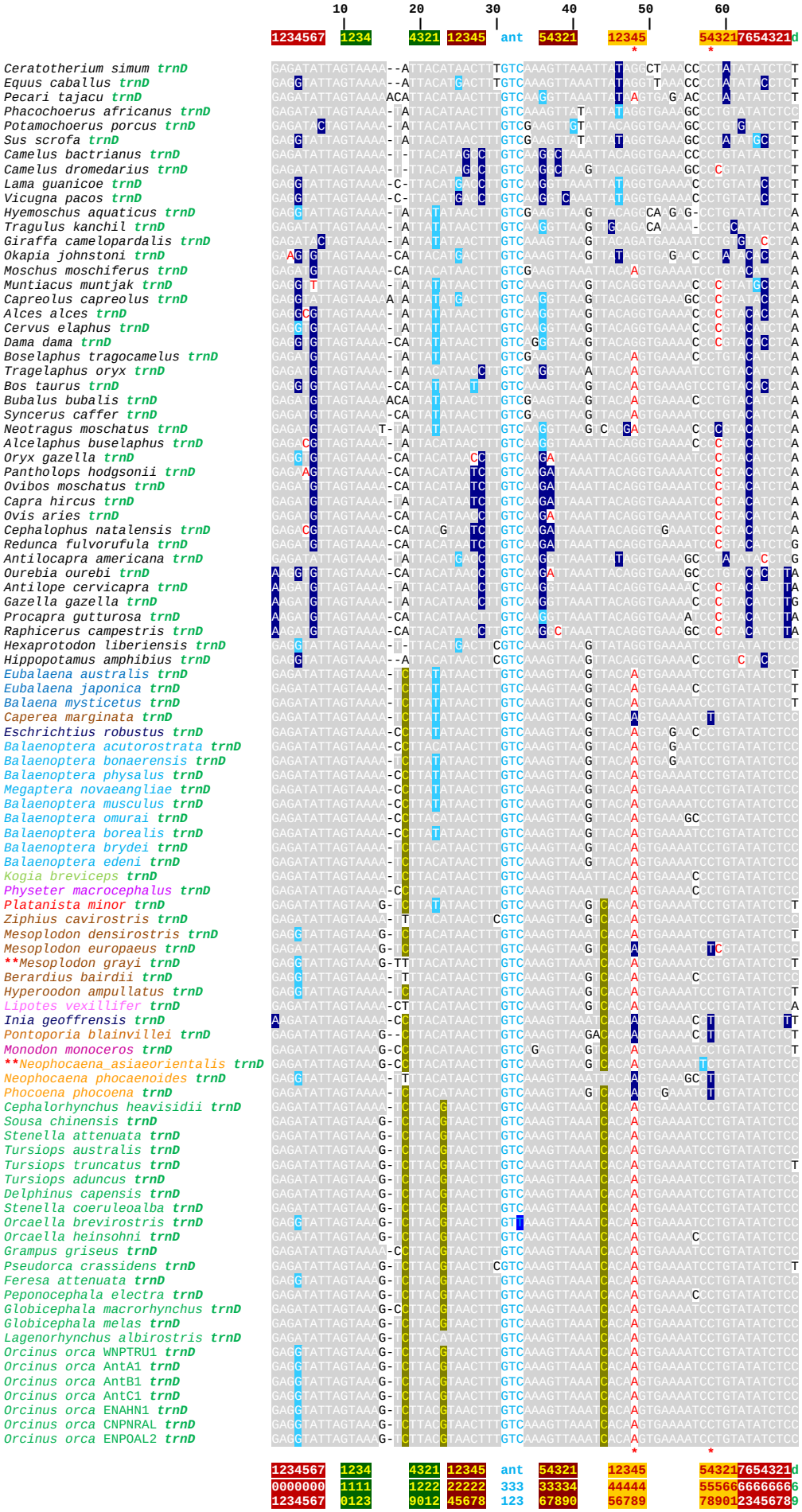
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T)
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent; N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-4(5) in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

***trnC* (CYS) multiple alignment**



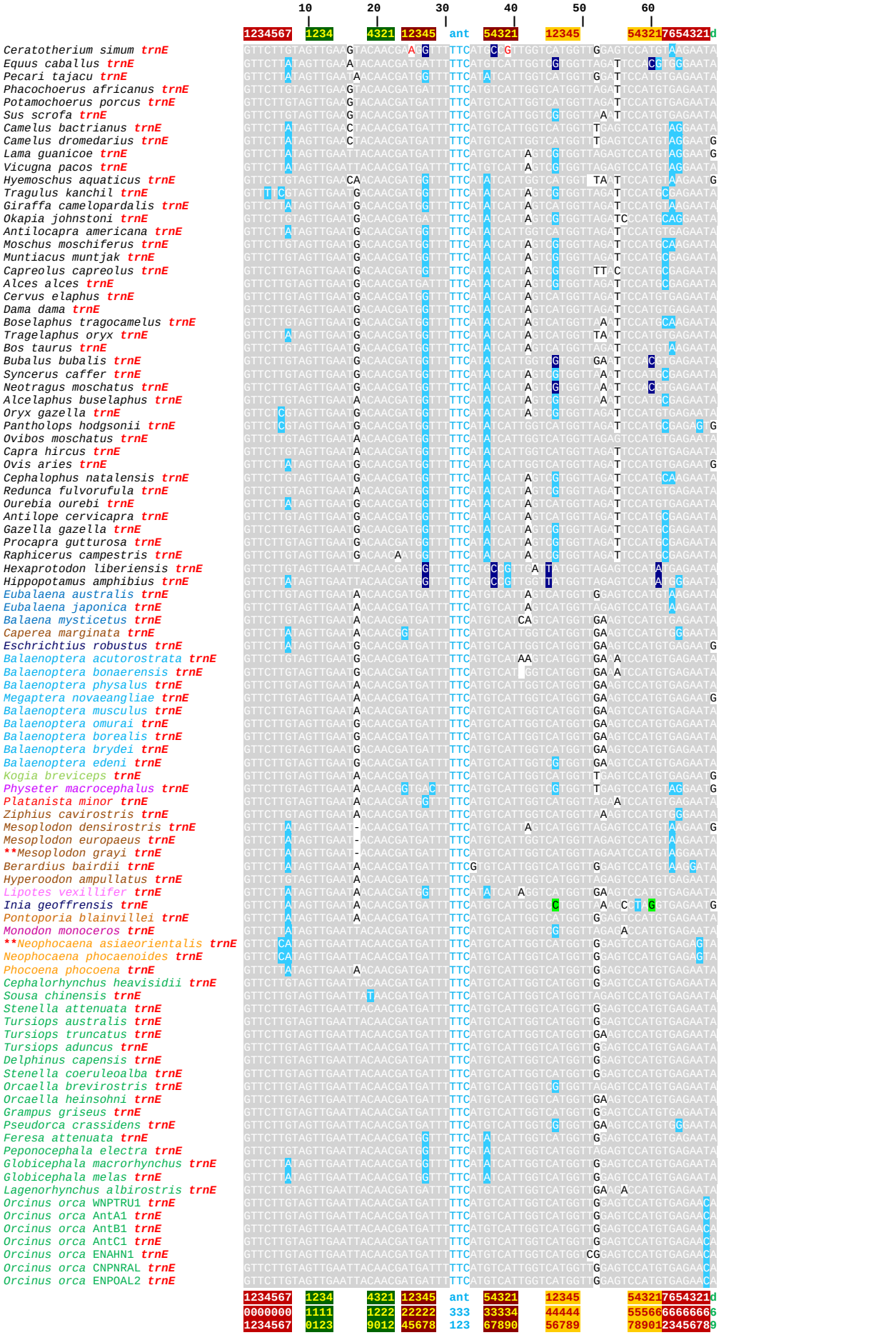
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-4 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnD (ASP) multiple alignment



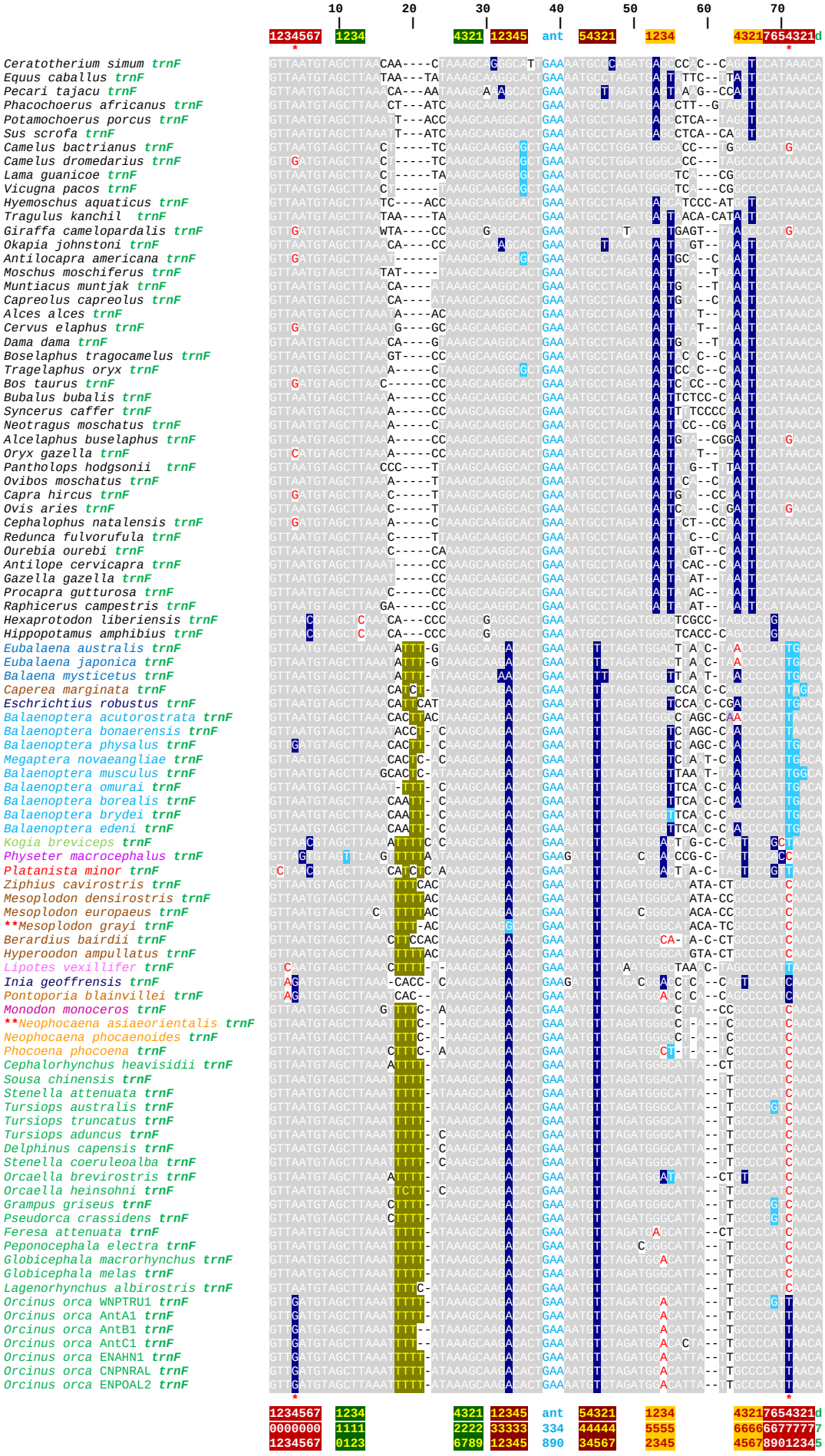
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-5 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnE (GLU) multiple alignment



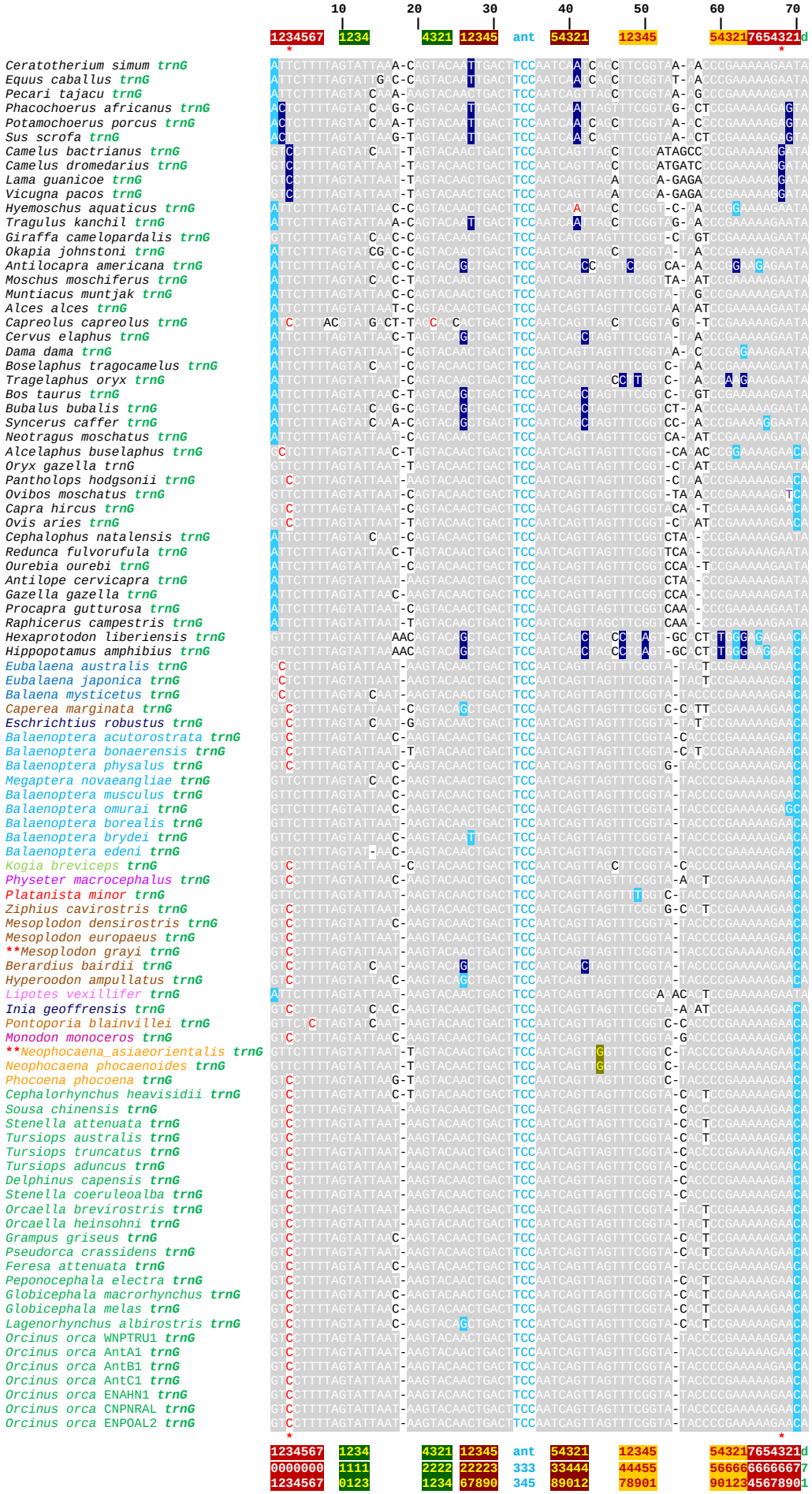
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-5 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnF (PHE) multiple alignment



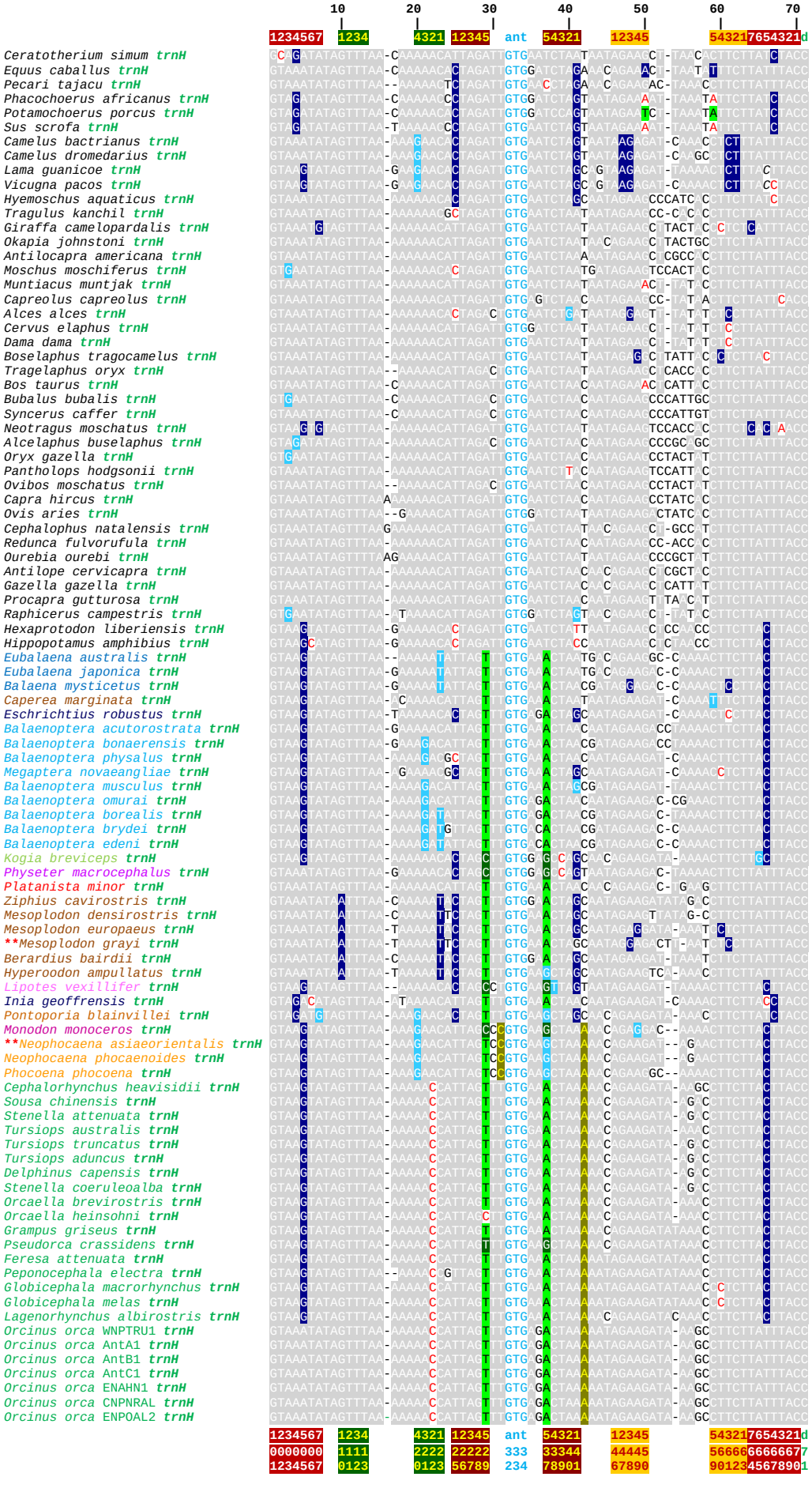
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-4 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnG (GLY) multiple alignment



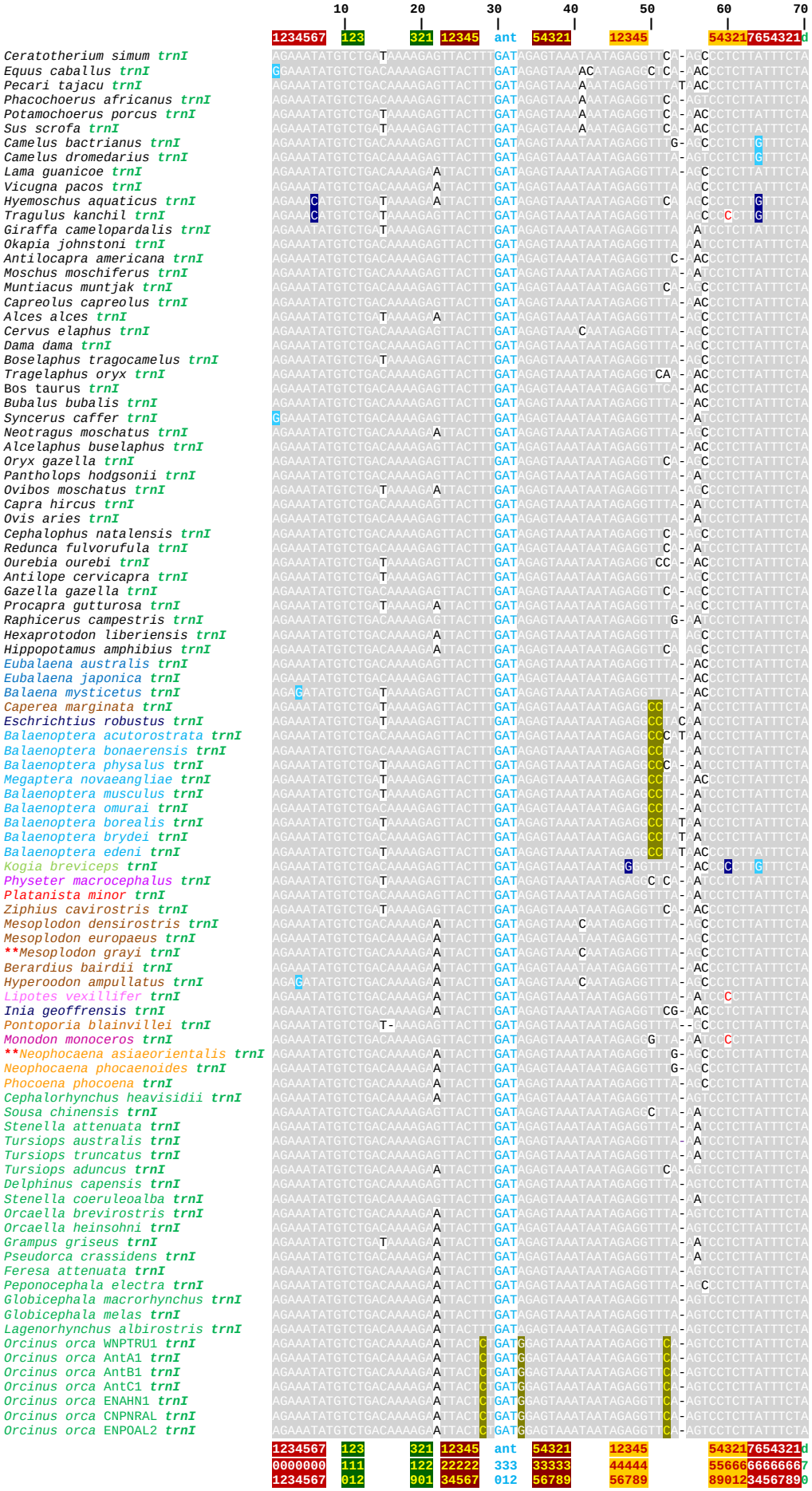
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N,N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N,N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-5 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnH (HIS) multiple alignment



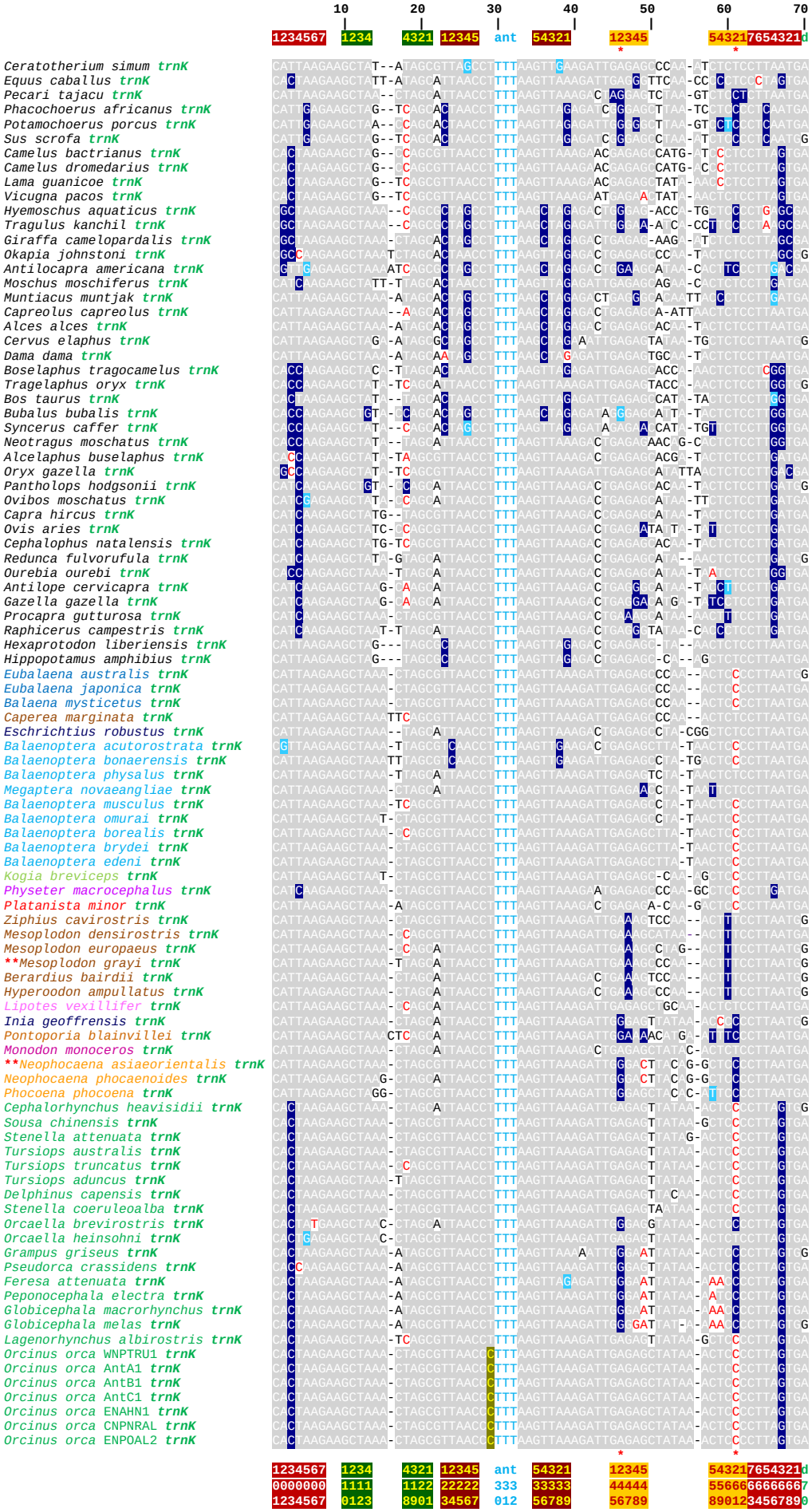
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T)
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-5 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnI (ILE) multiple alignment



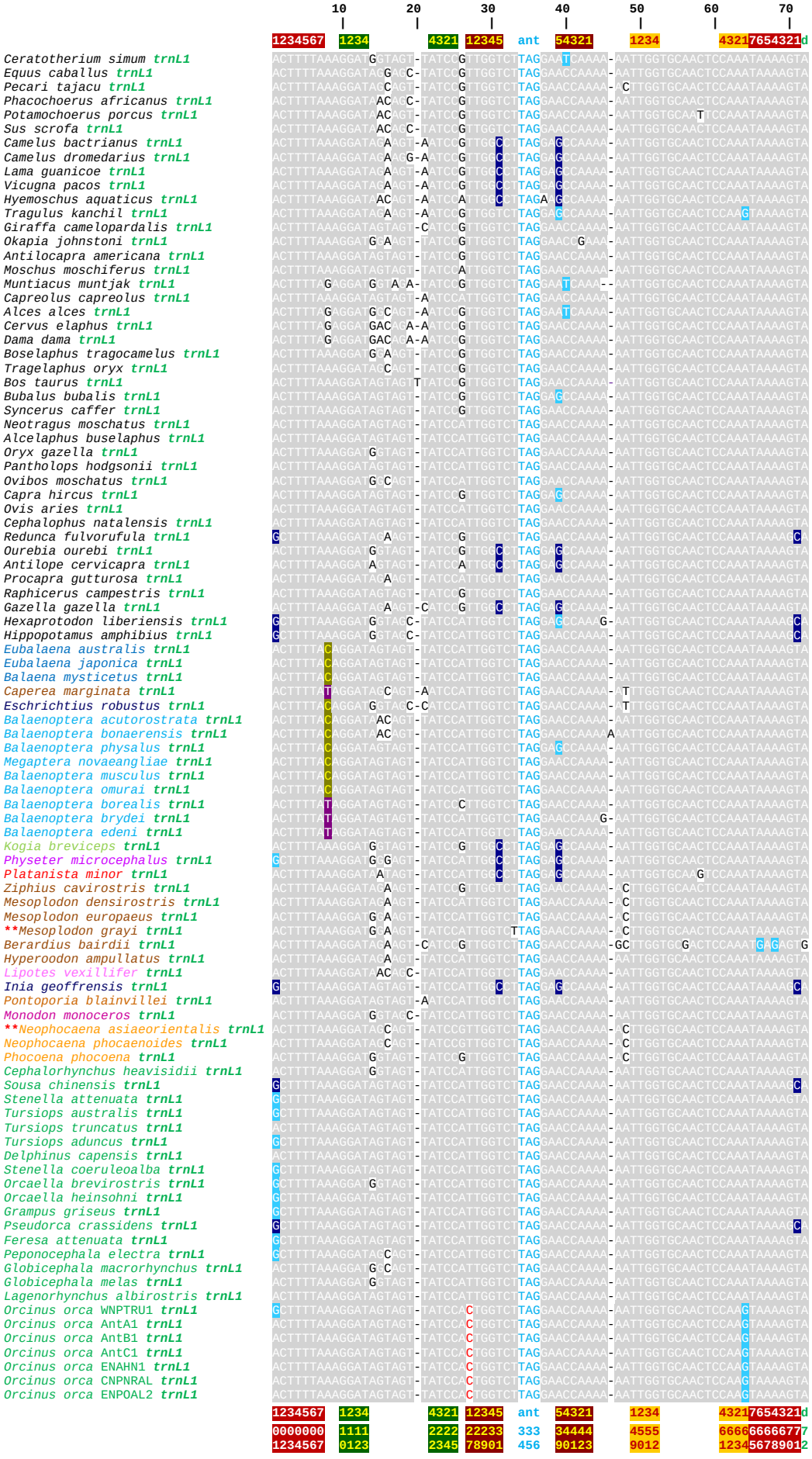
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N: molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-3 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-5 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnK (LYS) multiple alignment



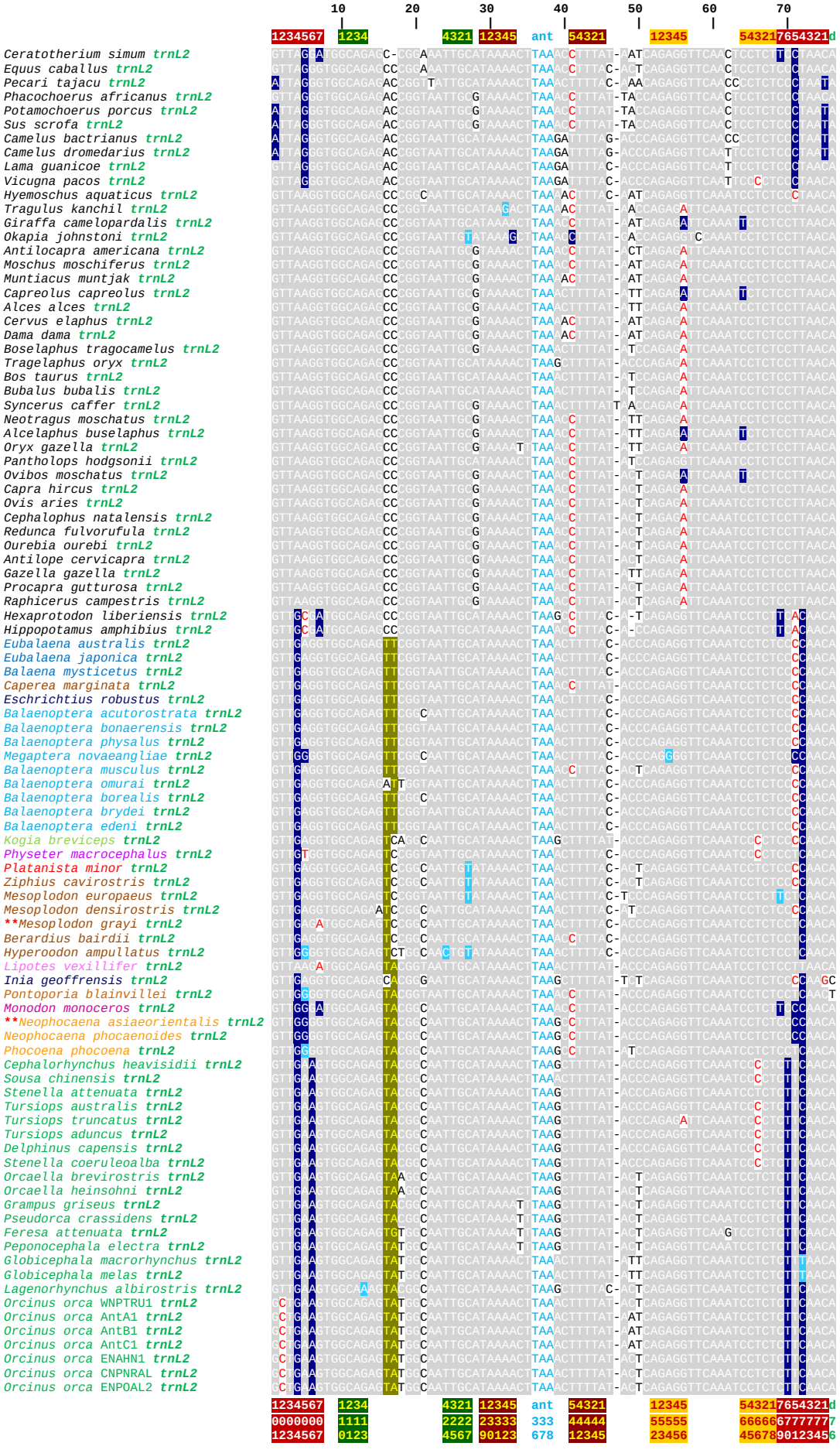
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T)
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-5 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnL1 [LEU (L1, CUN)] multiple alignment



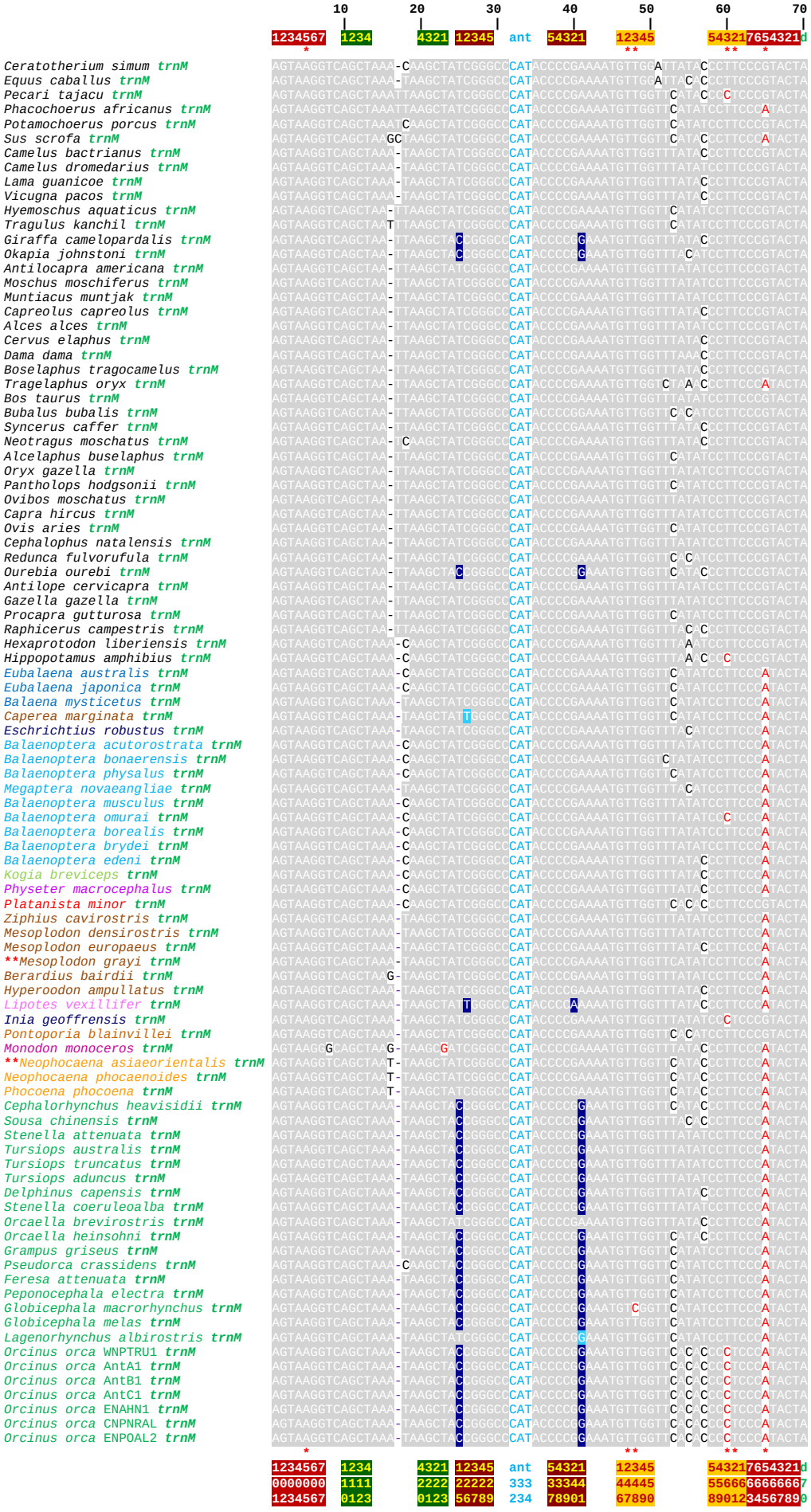
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-4 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnL2 [LEU (L2, UUR)] multiple alignment



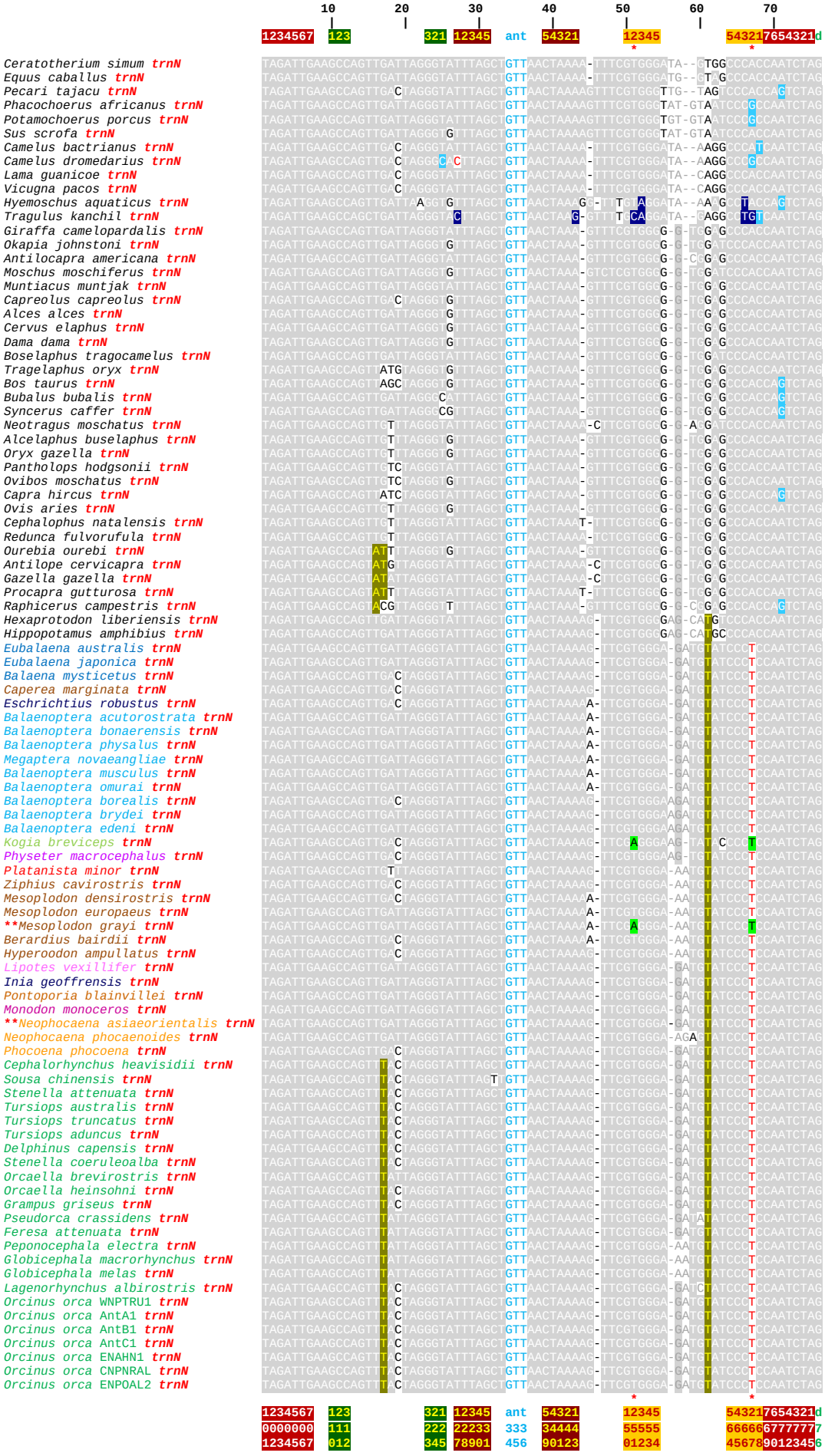
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-5 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnM (MET) multiple alignment



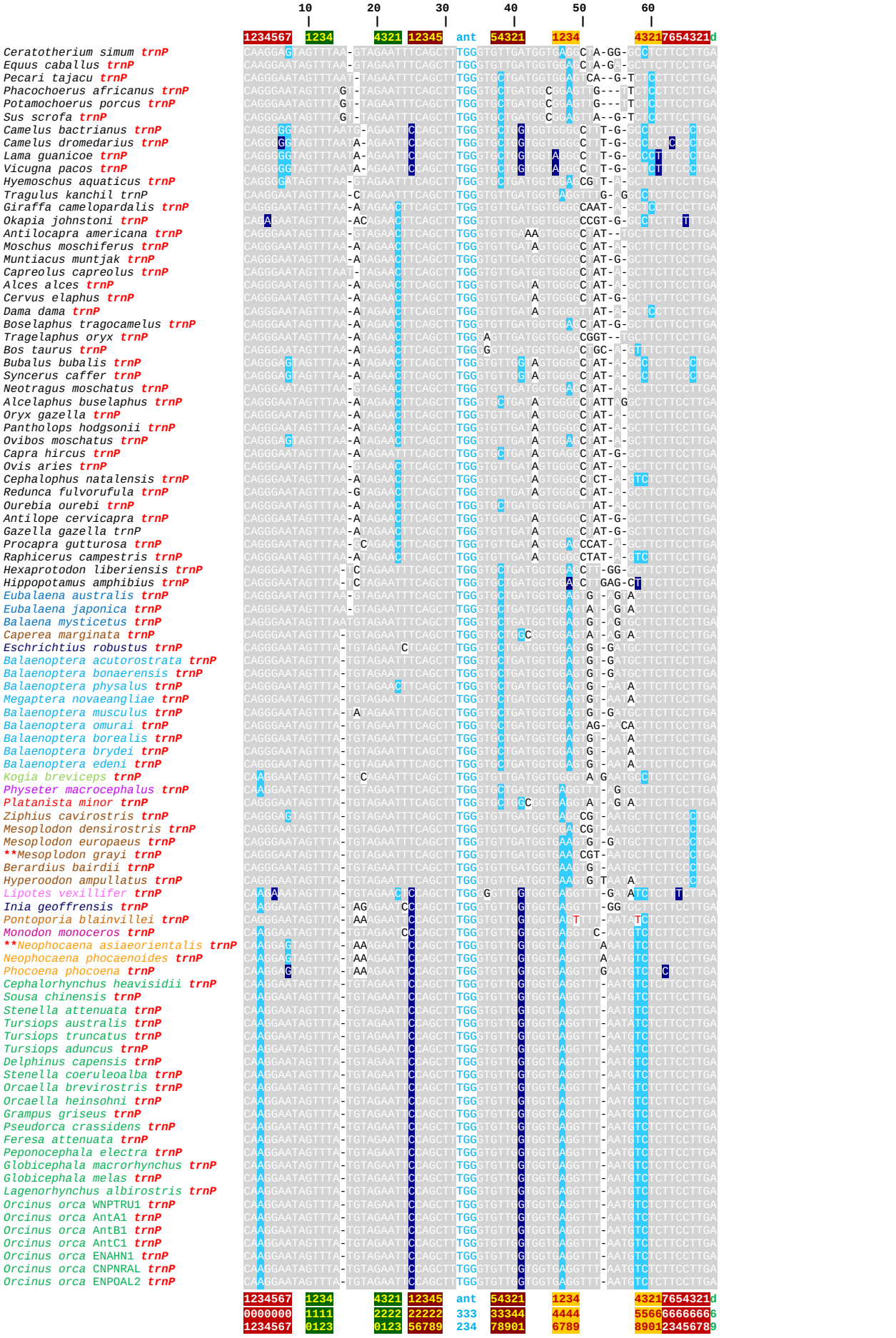
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N,N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N,N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-5 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnN (ASN) multiple alignment



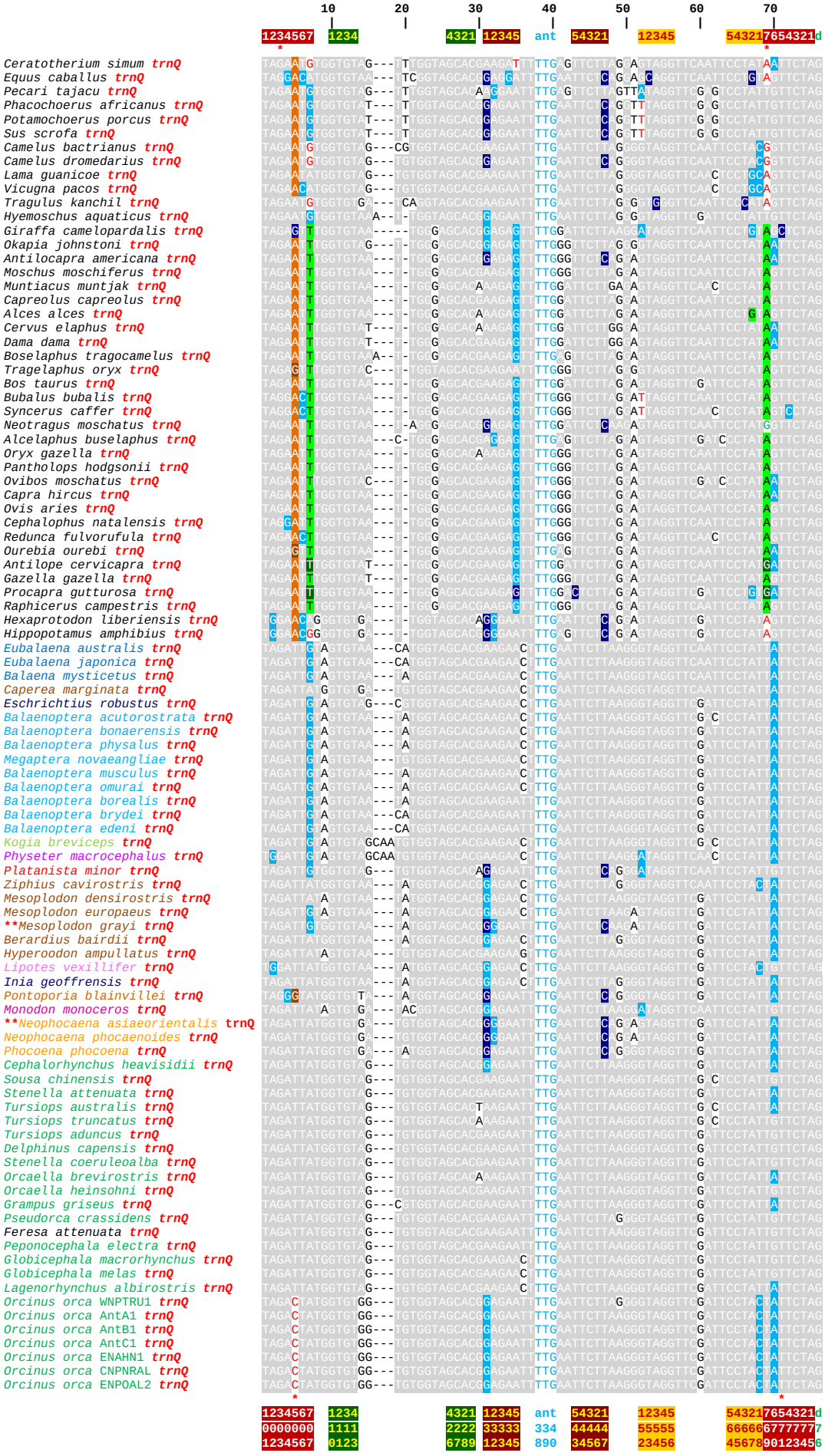
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-3 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-5 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

***trnP* (PRO) multiple alignment**



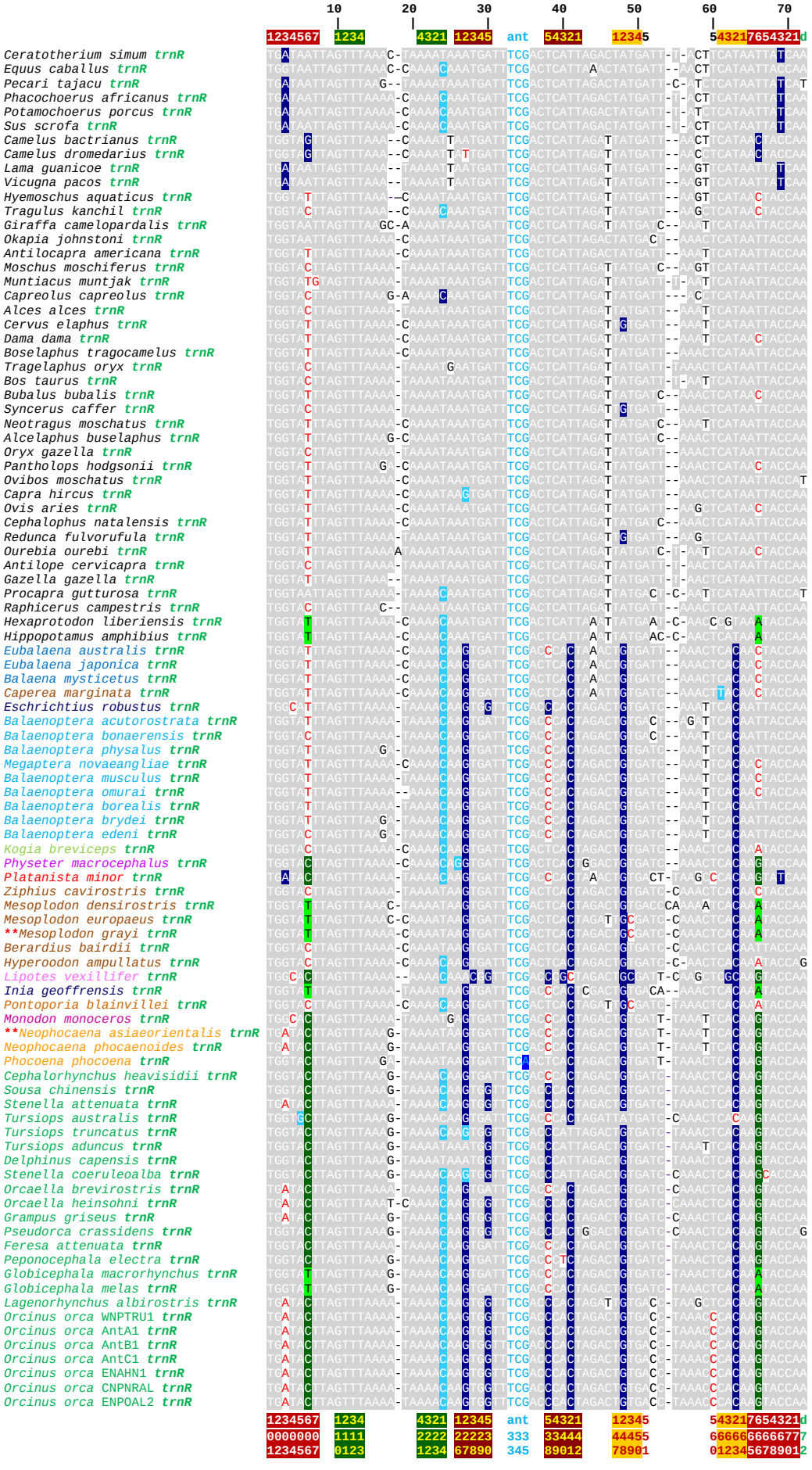
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-4 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnQ (GLN) multiple alignment



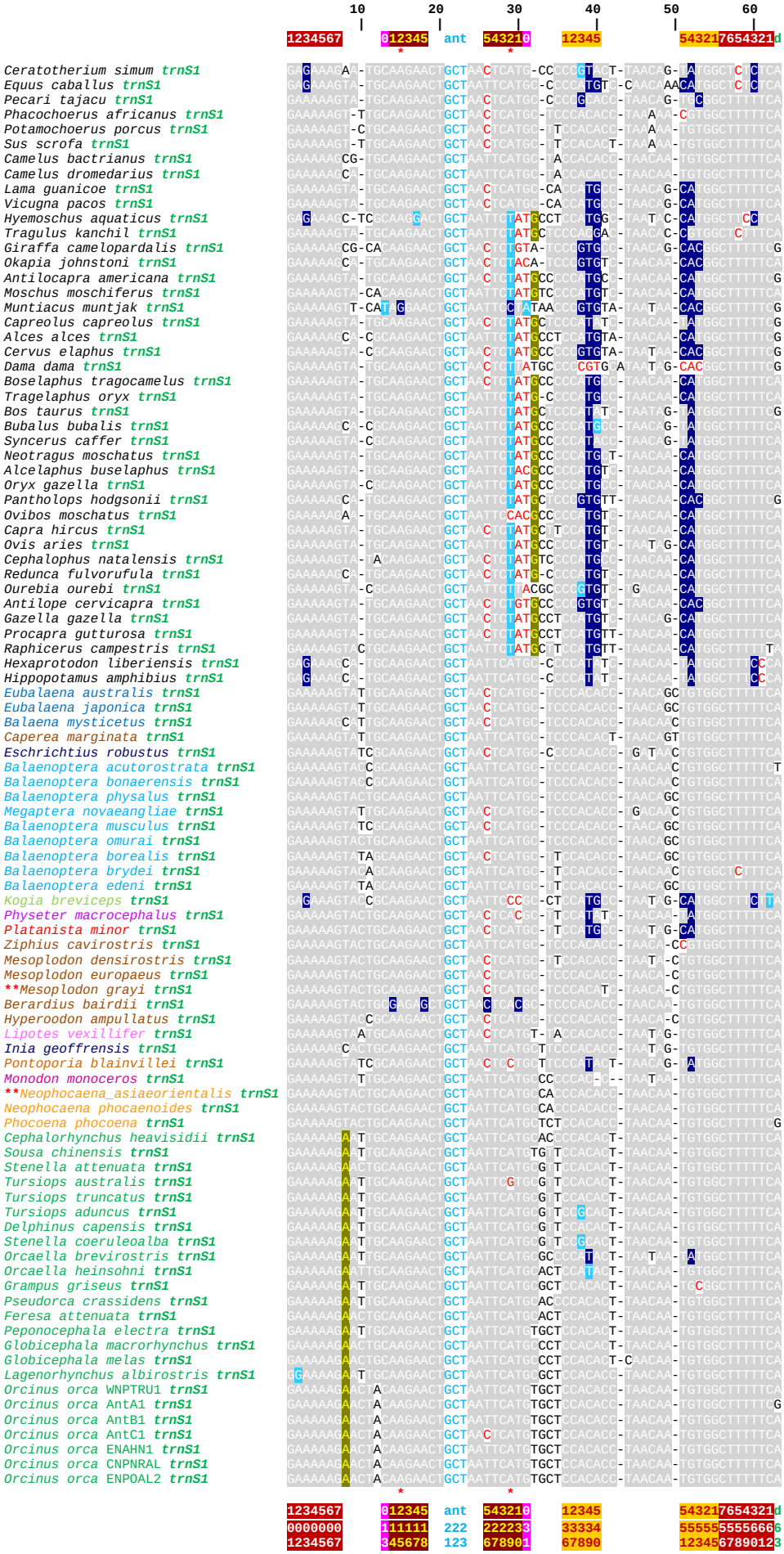
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N,N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N,N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-5 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnR (ARG) multiple alignment



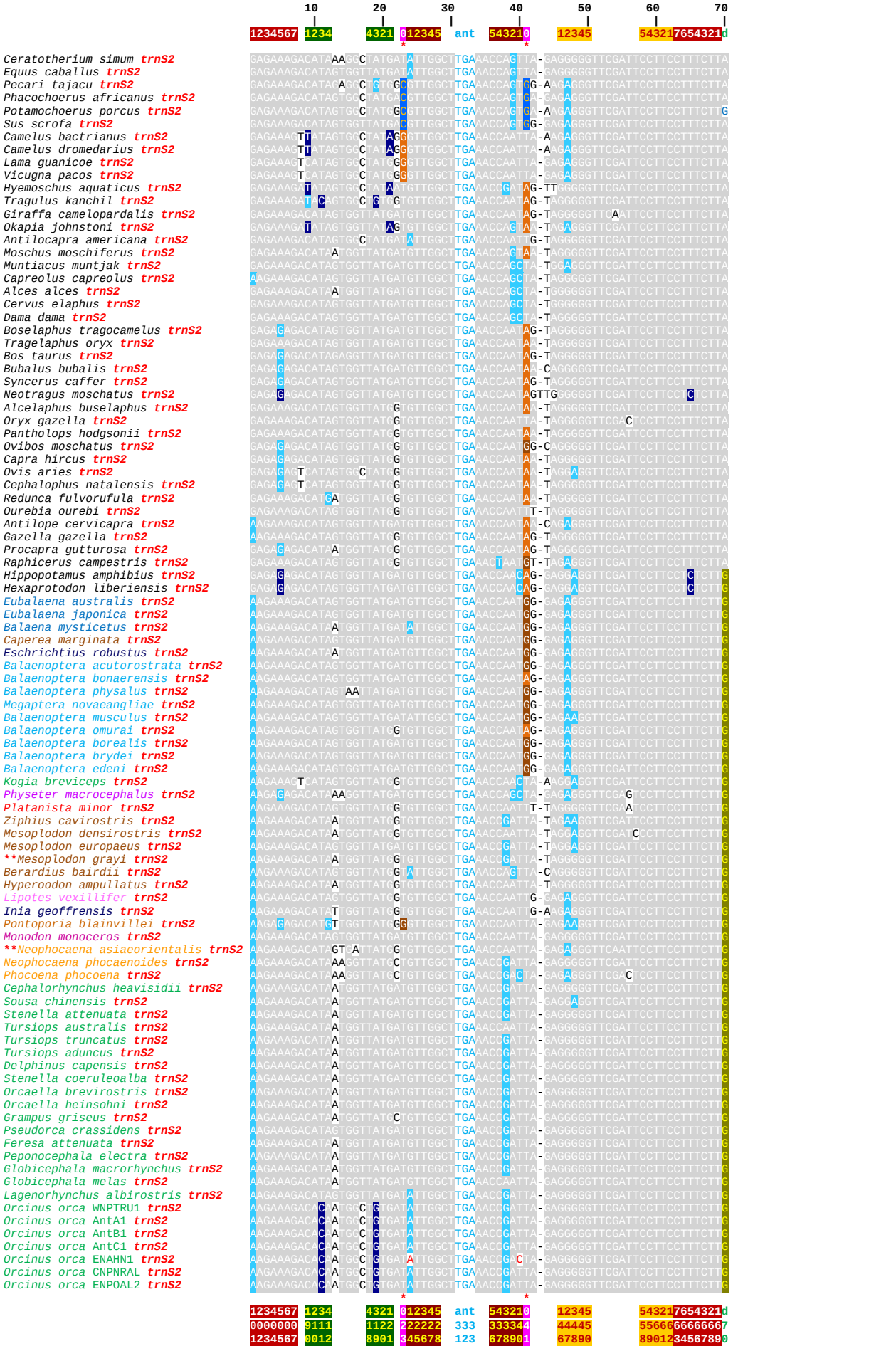
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-4(5) in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnS1 [SER (S1, AGY)] multiple alignment



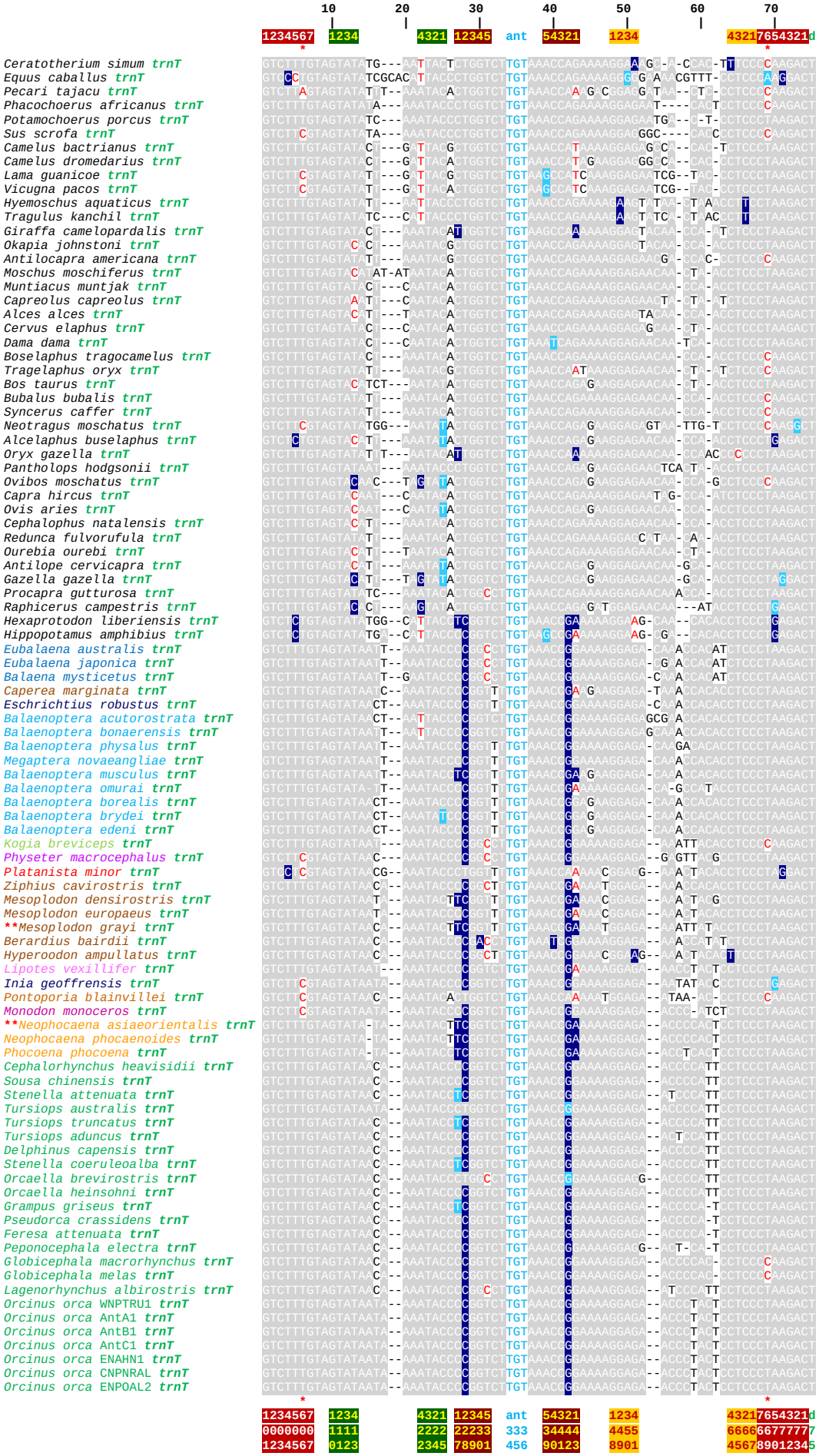
N, the most common base for the position.
N, N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; XX position (0)1-5 in the anticodon stem; X position 1-5 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnS2 [SER (S2, UCN)] multiple alignment



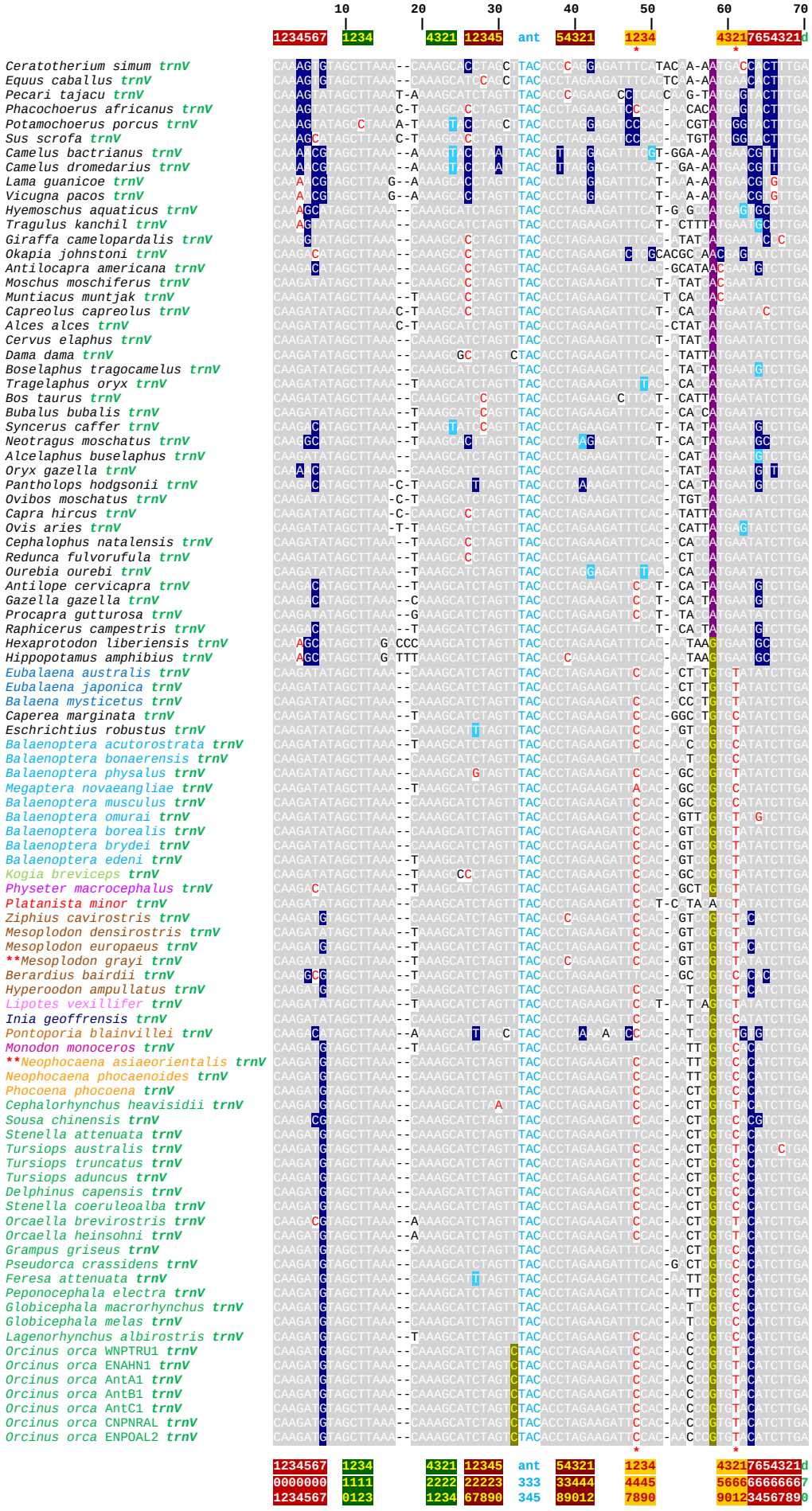
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N,N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N,N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a taxon.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; XX position (N)1-5 in the anticodon stem; X position 1-5 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnT (THR) multiple alignment



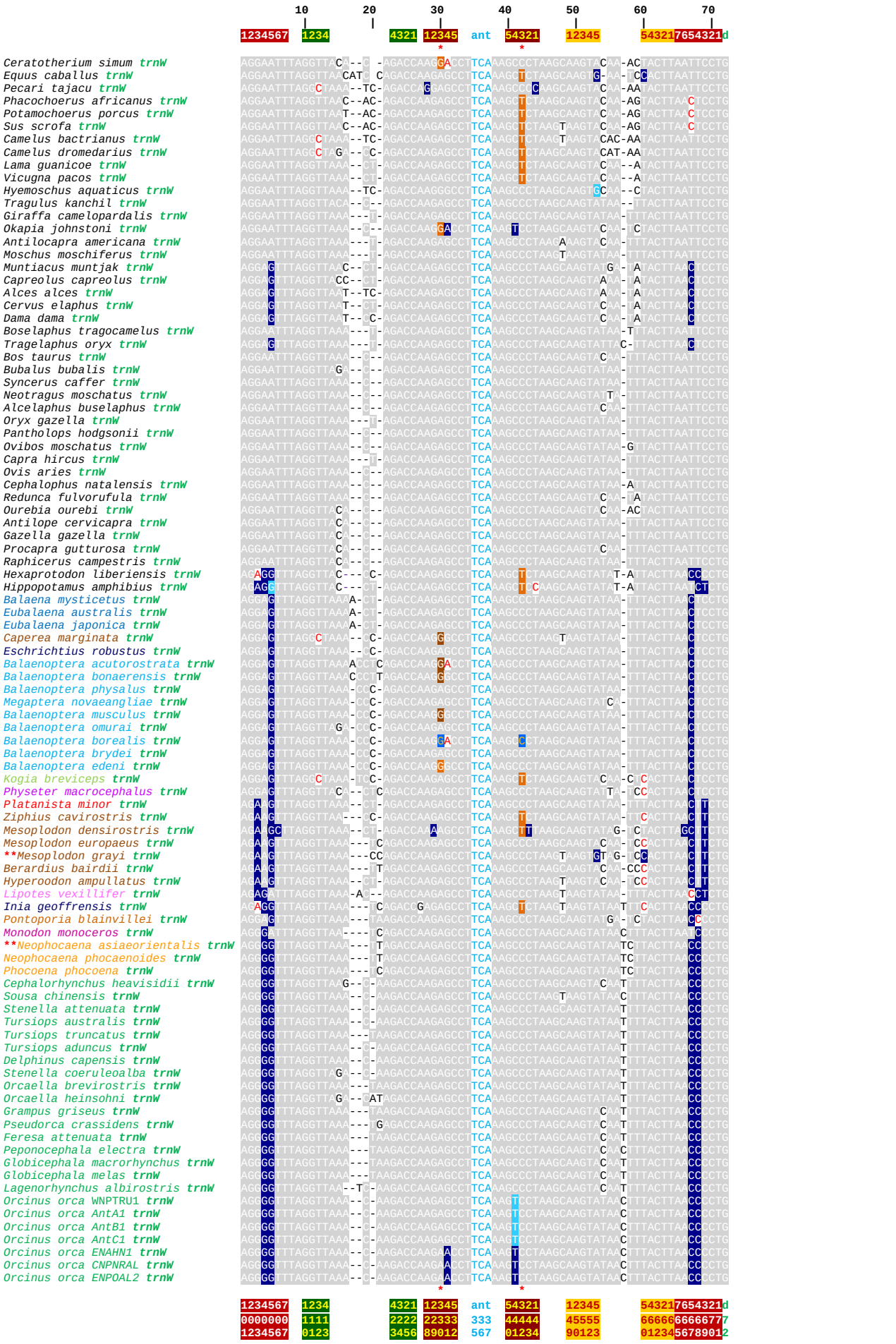
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a determined taxa.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-4 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnV (VAL) multiple alignment



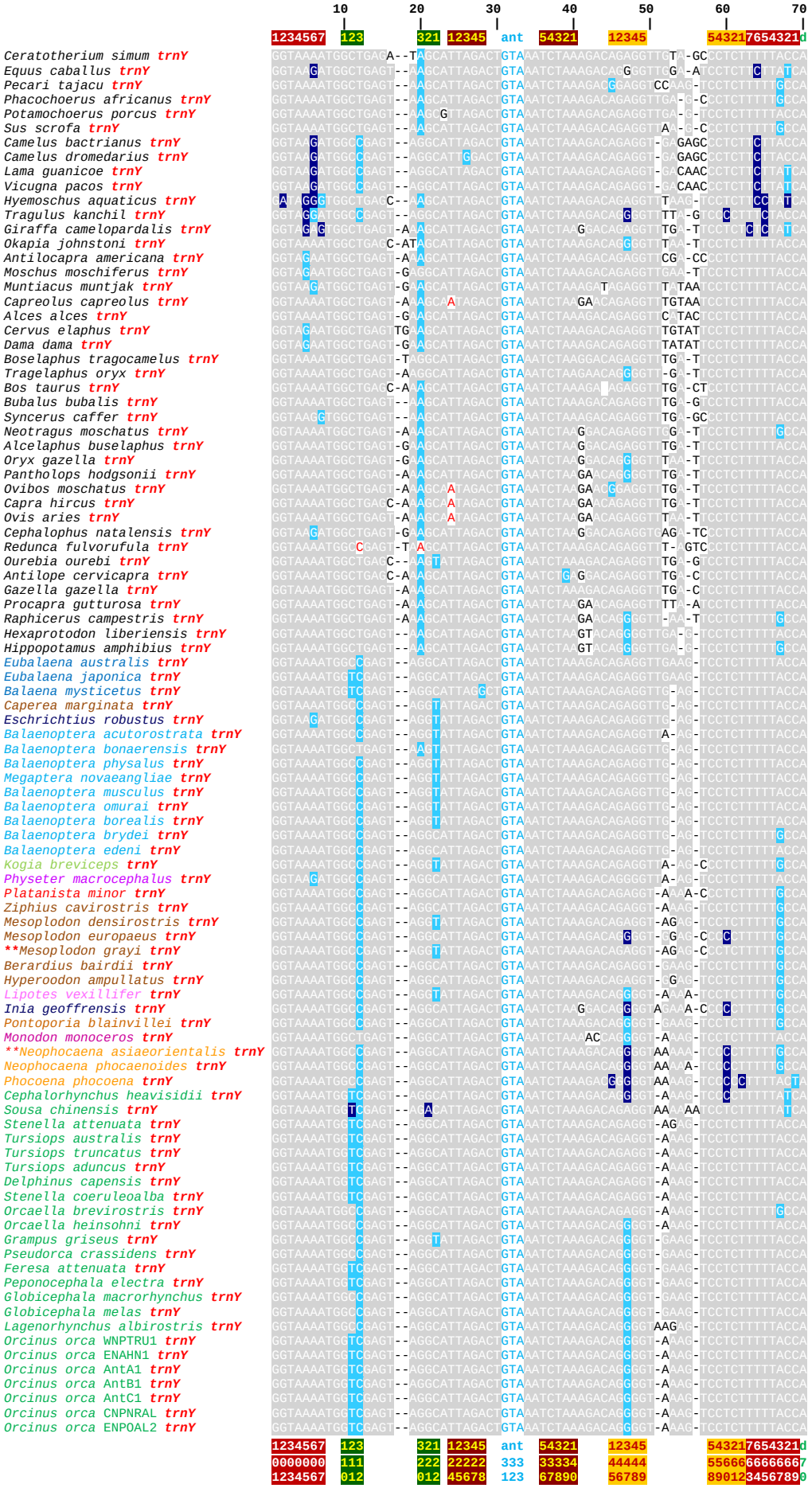
N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a determined taxa.
X, position 1-7 in the acceptor stem; X position 1-3 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-4 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnW (TRP) multiple alignment



N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a determined taxa.
X, position 1-7 in the acceptor stem; X position 1-4 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-5 in the TΨC stem; ant, anticodon; d, discriminator nucleotide.

trnY (**TYR**) multiple alignment



N, the most common base for the position.
N, half compensatory base change in the stem pair (e.g. T – G vs C – G; A-T vs G-T).
N, N, half compensatory base change in the stem pair exhibiting a mismatch (e.g. T–A vs A–A). Different colours are used to better differentiate the changes.
N, fully compensatory base change in the stem pair exhibiting a mismatch (e.g. C–G vs T–T).
N, type I fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs purine – pyrimidine, e.g. G – C vs A – T).
N, N, type II fully compensatory base change in the stem pair (i.e. purine – pyrimidine vs pyrimidine – purine, e.g. A – T vs T – A). Different colours are used to better differentiate the changes.
N, a mismatch in the in the stem pair; N, substitution pattern not modelled; *, pair in the stem in which a mismatch is prominent. N, N, molecular signature for a determined taxa.
X, position 1-7 in the acceptor stem; X position 1-3 in the DHU stem; X position 1-5 in the anticodon stem; X position 1-5 in the TΨC stem; **ant**, anticodon; **d**, discriminator nucleotide.