

Comparative Analysis of Mitochondrial Genomes

1. Introduction

An analysis was performed on the 13 protein-coding mtDNA regions in several groups of metazoan species in order to:

- detect patterns of gene-specific mitochondrial substitution rate signatures,
- explore the dependencies of the base pair composition with the nucleotide substitution rate and
- detect any group-specific differences.

A total of 869 complete mitochondrial genomes were downloaded from GenBank 179 (release December 2006). Data were grouped in closely related groups (i.e. *genus*). For each such group, coding regions of the 13 individual genes were extracted, aligned [1] and used to compute the rate of nonsynonymous (dN) to synonymous (dS) changes within groups [2] and the base pair composition (ATs).

The datasets provide fairly broad representation of vertebrates taxonomies (74 genera), but the invertebrates are under-represented (20 genera) (Table 1).

Vertebrates		Invertebrates	
Classification	Datasets	Classification	Datasets
Amphibia	10(30)	Insecta	4(11)
Mammalia	23(71)	Mollusca	3(8)
Sauria	10(27)	Cnidaria	2(5)
Teleostei	25(76)	Crustacea	4(8)
Testudines	2(6)	Chelicerata	3(9)
Chondrichthyes	1(2)	Platyhelminthes	2(5)
Acipenseriformes	2(5)	Echinodermata	1(2)
Semionotiformes	1(3)	Branchiostoma	1(4)
Total	74(220)		20(52)

Table 1. Classification and distribution of data sets. The numbers in parenthesis show the total number of species used in the classification (e.g. 30 amphibian species are grouped in 10 genera with an average of 3 species per group).

2. Results

Nonsynonymous substitutions

◆ ATP8: the largest dN variability

Vert: $\min_{dN}=0.009$, $\max_{dN}=0.165$, $\text{stdev}_{dN}=0.060$;

Invert: $\min_{dN}=0$, $\max_{dN}=0.303$, $\text{stdev}_{dN}=0.113$

◆ COX1: the least dN variability

Vert: $\min_{dN}=0.001$, $\max_{dN}=0.026$, $\text{stdev}_{dN}=0.008$;

Invert: $\min_{dN}=0$, $\max_{dN}=0.051$, $\text{stdev}_{dN}=0.019$.

◆ the invertebrates show a much broader range of nonsynonymous rates (Invert_ $\text{stdev}_{dN} = 0.037$) than the vertebrates (Vert_ $\text{stdev}_{dN} = 0.015$) (Fig. 1).

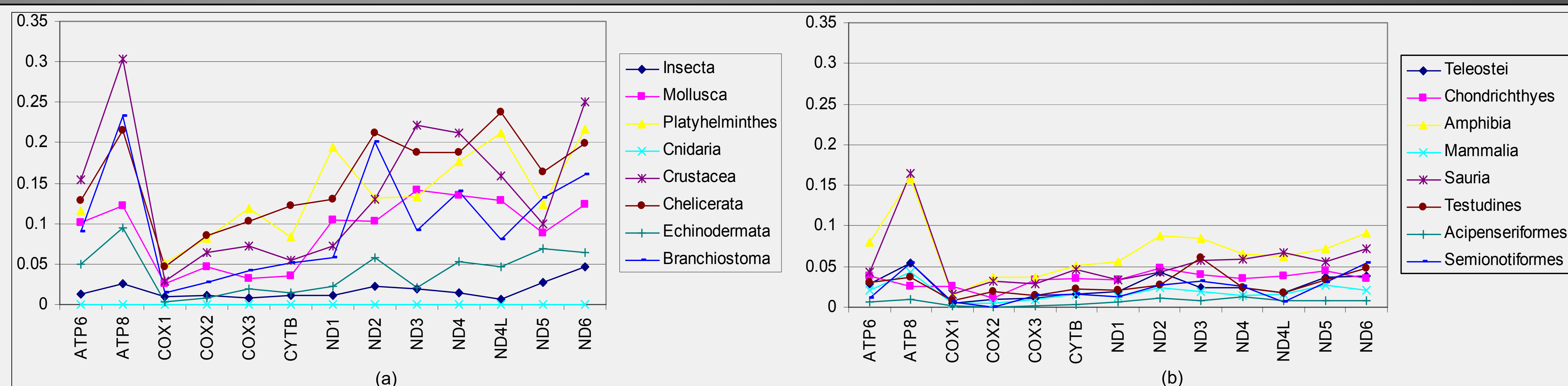


Figure 1. Nonsynonymous substitutions rate for (a) invertebrates and (b) vertebrates

Synonymous substitutions

◆ ATP8: low dS variability

Vert: $\min_{dS}=0.050$, $\max_{dS}=0.555$, $\text{stdev}_{dS}=0.181$;

Invert: $\min_{dS}=0$, $\max_{dS}=1.381$, $\text{stdev}_{dS}=0.508$.

◆ Dramatic difference in the variability of dS estimates between invertebrates (Invert_ $\text{stdev}_{dS} = 0.377$) and vertebrates (Vert_ $\text{stdev}_{dS} = 0.092$) (Fig. 2).

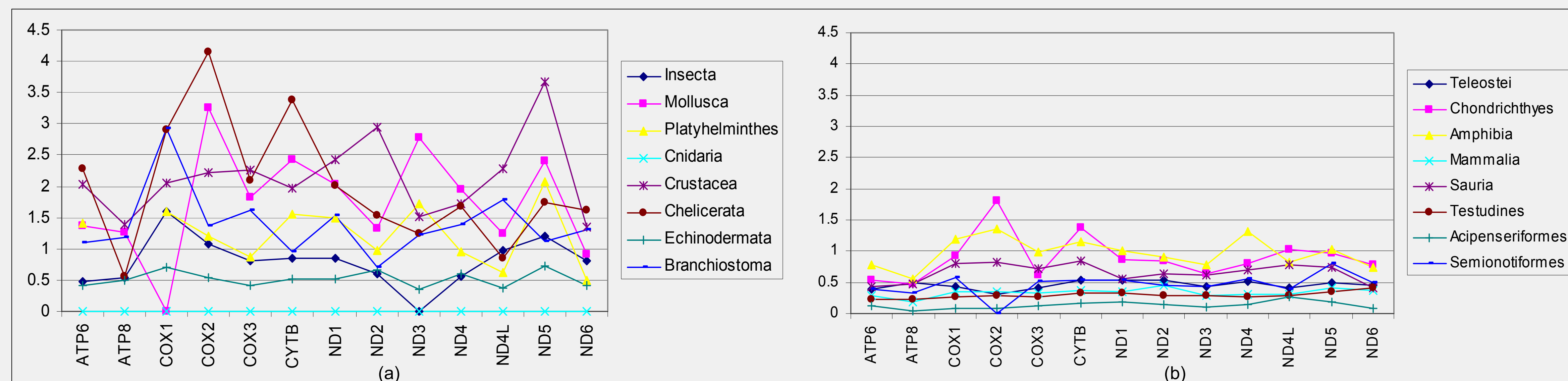


Figure 2. Synonymous substitutions rate for (a) invertebrates and (b) vertebrates

AT base pair concentration

◆ higher AT concentration of the coding sequences in invertebrates (Invert_ $\text{AT} = 0.669$) than in vertebrates (Vert_ $\text{AT} = 0.569$) (Fig. 3.a).

◆ ATP8: the highest AT base pair concentration for both vertebrates (Vert_ $\text{avg}_{AT} = 0.605$) [3] and invertebrates (Invert_ $\text{avg}_{AT} = 0.735$)

◆ Correlation with dN values: Vert_ $r = 0.717$, Invert_ $r = 0.916$ (Fig. 3.b)

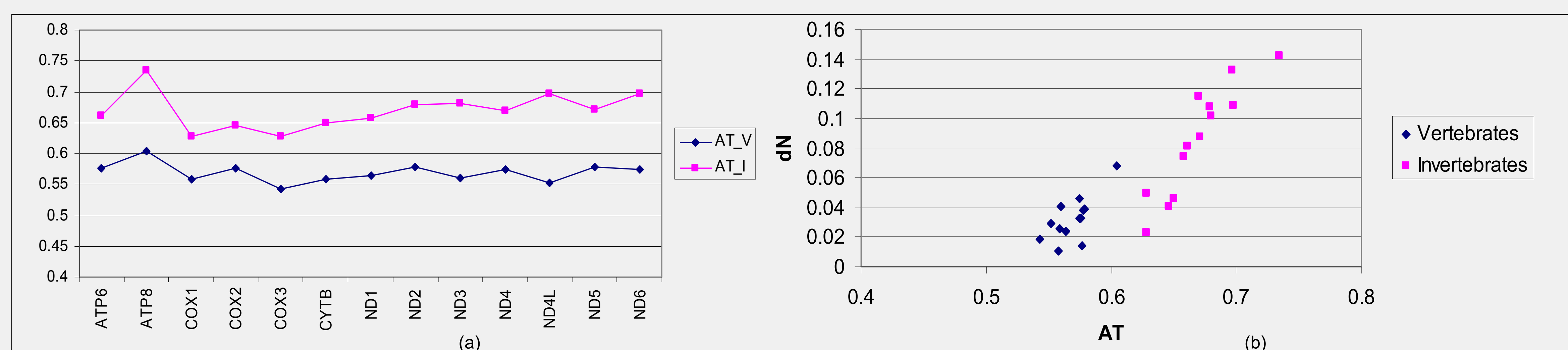


Figure 3. (a) AT base pair concentration in vertebrates (blue) and invertebrates (pink) and (b) correlation between AT base pair concentration and dN values for vertebrates (blue) and invertebrates (pink)

The influence of the amino-acids frequency of occurrence was tested by recording the presence of each amino-acid in each gene from the mitochondrial genomes (grouped on vertebrate / invertebrate classification).

Phe, Val, Ser, Tyr, and Gly are present in a higher concentration in invertebrates than in vertebrates, while Pro, Thr, Leu and Ala have a higher concentration in vertebrates than in invertebrates.

Further work is needed to identify the causes of the amino-acid frequency discrepancies and comparisons with nuclear DNA genomes should provide more insights.

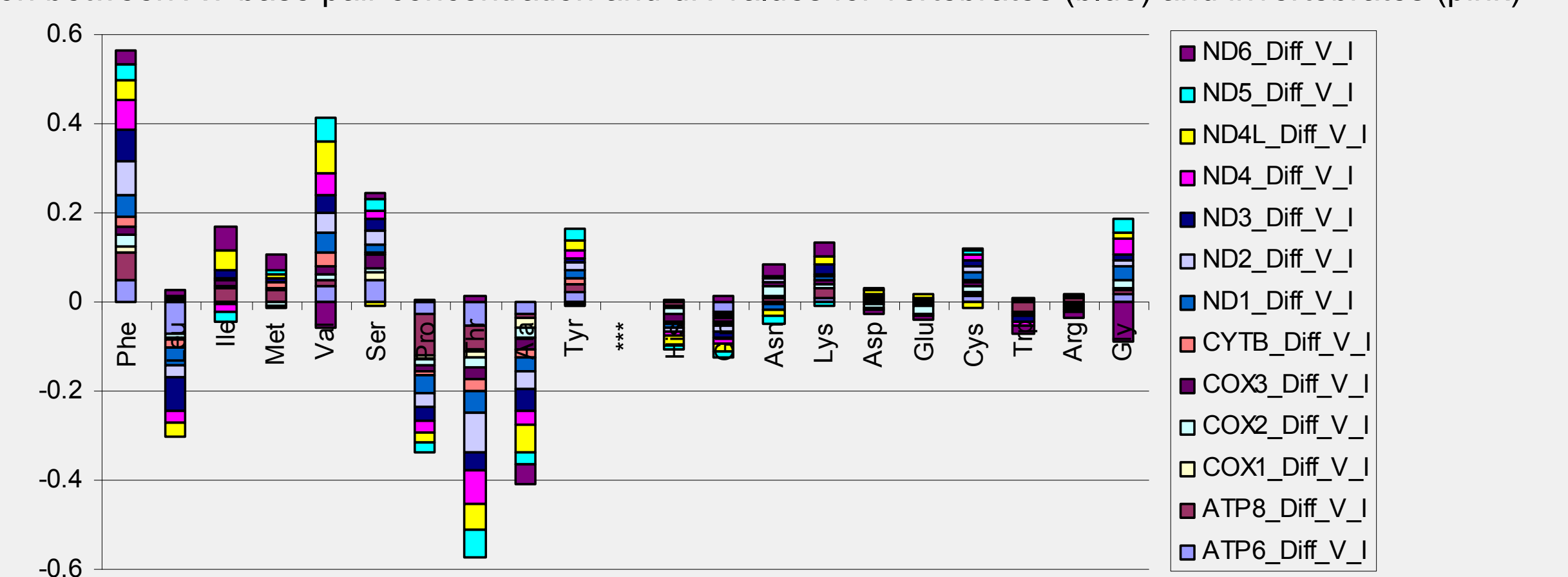


Figure 4. Differences in amino acids percentages between invertebrates and vertebrates in each CDS

3. Conclusions

The current analysis highlights important signals that need to be considered. Each protein coding gene in mtDNA evolve differently (the ATP8 gene, missing for Nematodes and Platyhelminthes, [5] has a high variability, while the COX1 gene is the least affected by nonsynonymous substitutions). Regarding the species classifications, Cnidaria is the slowest evolving invertebrates group [4], while the Amphibia is the fastest evolving vertebrates group.

Careful analysis must be implemented when dealing with dN/dS ratio. The dN/dS ratio for Cnidaria analyzed in the ND4 gene has a high value (0.34315) but this is only a result of a fraction of very low and "similar" values of dN (0.00058) and dS (0.0017). The correct conclusion of this result is the low variability of Cnidaria group which is in contrast with the one obtained if only the dN/dS fraction value is considered. Our results show a different pattern of evolution in vertebrates (low dN, dS) than in invertebrates (high dN, dS) while the dN/dS ratios provide little information. Moreover, our results do not confirm the dS values obtained by Bazin et al. [6], and most of their interpretations are based on dN/dS ratios. Further research is needed, including similar analyses on nuclear DNA genes.

References

- Thompson J.D. et. al. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice, *Nucleic Acids Research*, 22, 4673-4680.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood, *Computer Applications in BioSciences*, 13, 555-556.
- Pesole G. et. al. Nucleotide substitution rate of mammalian mitochondrial genomes, *J Mol Evol.*, 48 (4), 427-34.
- Shearer T. L. et. al. Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria), *Molecular Ecology*, 11(12), 2475-2487.
- Vallès Y., Boore J. L., Lophotrochozoan mitochondrial genomes, *Integrative and Comparative Biology*, 46 (4), 544-557.
- Bazin E. et. al. Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals, *Science* 312 (5773), 570.