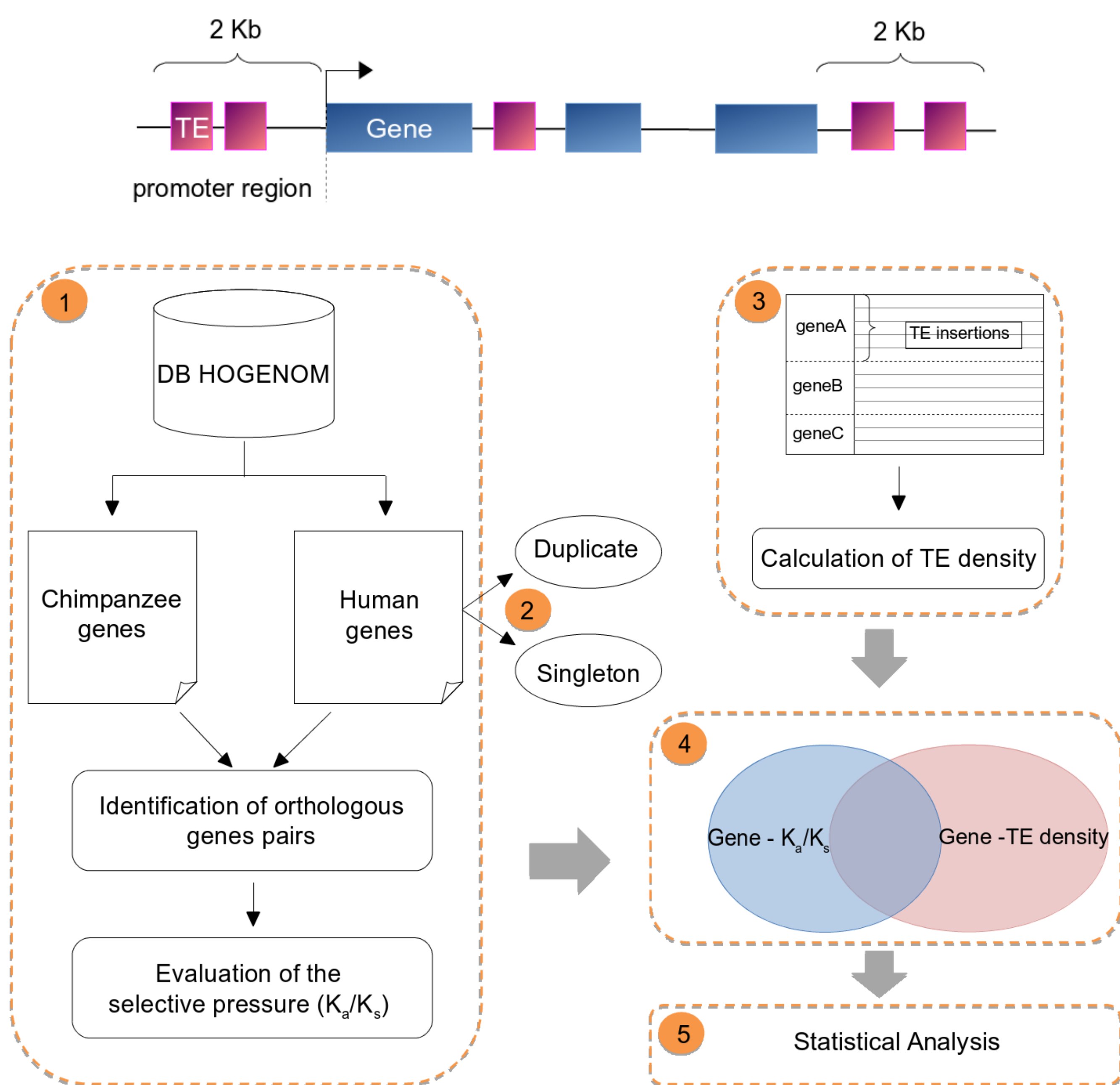


## INTRODUCTION & OBJECTIVE

It is widely agreed that gene duplication is an engine for the apparition of new gene functions (Ohno, 1970). The fate of most **duplicate genes** is the loss of function (Semon et Wolfe, 2007; Vinogradov, 2012), but a significant part of duplicate copies remain functional. The evolutionary mechanisms explaining the duplicate genes maintenance are still poorly understood. **Transposable elements (TEs)** are repeated genomic sequences which participate to genome evolution. They constitute 55% of the human genome (Cordaux et Batzer, 2009). TEs are also known to influence the expression of nearby genes. Here we explore the relationship between TE context and gene type (duplicate and singletons).

Is there a link between the TE context and duplicate and singleton genes? Is there a link between the selective pressure and duplicate and singleton genes?

## MATERIALS & METHODS



### 1 Evaluation of selective pressure between human and chimpanzee orthologous gene pairs ( $K_a/K_s$ )

- Construction of Human and Chimpanzee datasets from an alignment tree file downloaded from HOGENOM, a database of homologous genes

	Chimpanzee	Human
Genes	18390	19582
Gene families	11507	11630

- Identification of orthologous gene pairs with Best Reciprocal Hits (BRH) 16714 gene pairs
- Estimation of the rate of non-synonymous substitutions ( $K_a$ ) and the rate of synonymous substitutions ( $K_s$ ) by the yn00 (method of Yang and Nielsen, 2000) program of PAML (Phylogenetic Analysis by Maximum Likelihood)

### 2 Identification of duplicate genes and singletons

For a given homologous gene family, a gene is duplicated if there is more than two genes of a same species.  
Human 19582 genes, 11161 duplicate genes and 8421 singleton genes

### 3 Calculation of overall TE density ( $D_{ET}$ ) of 20806 genes

$$D_{ET} = \frac{\text{Number of TEs}}{\text{Gene length} + (2 \times 2Kb) - \text{TE length}}$$

### 4 Merging files

### 5 Statistical Analysis

Analysis of covariance, model equation :

$$Y = \mu + \alpha T + \beta P + \gamma TP + \epsilon$$

where Y is the TE density, T is the qualitative variable duplicate/singleton and P is the quantitative variable  $K_a/K_s$  ratio

## RESULTS

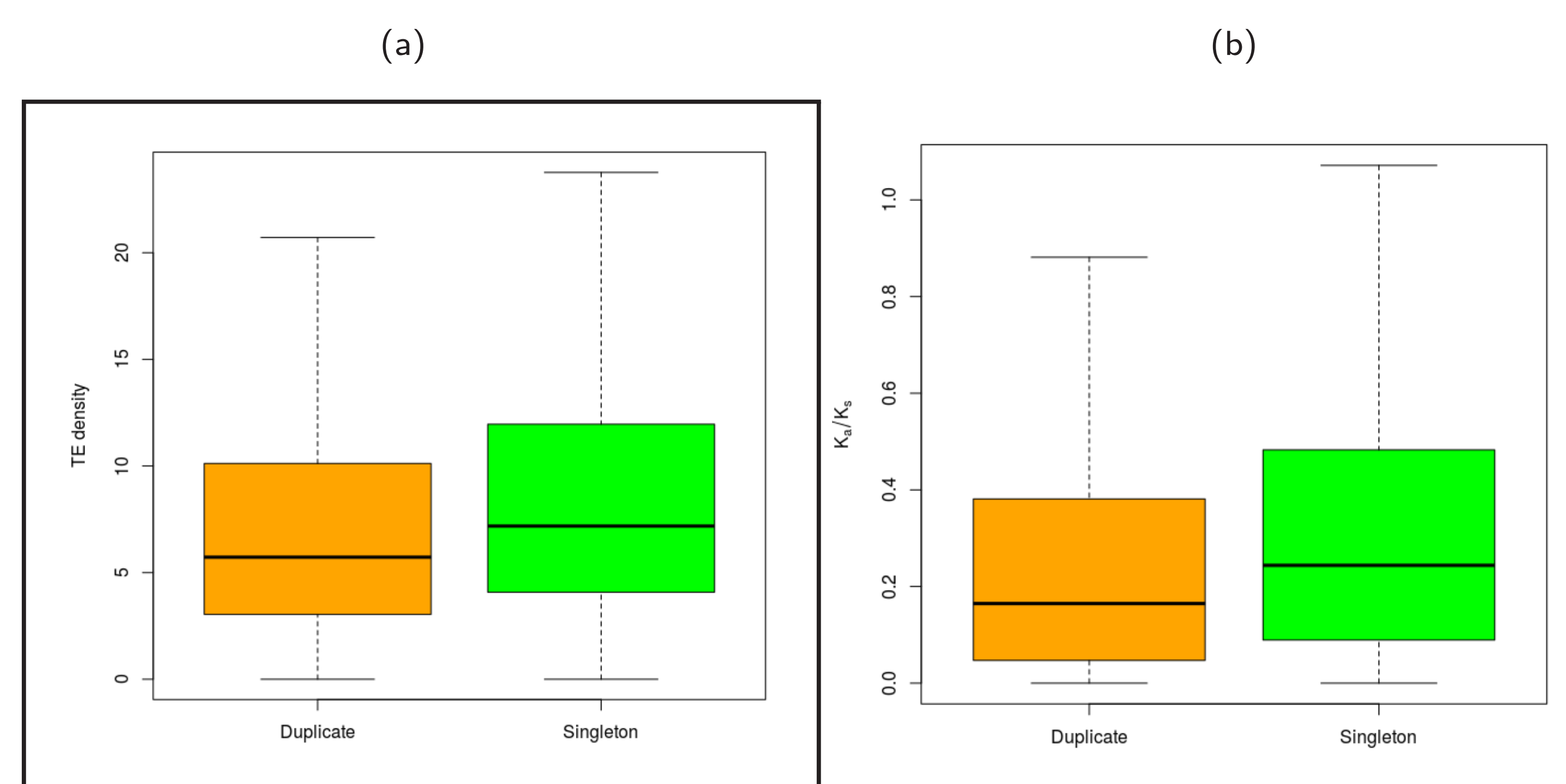


Figure 1: Density of TEs (a) and  $K_a/K_s$  ratio (b) according to the duplicate and singleton genes

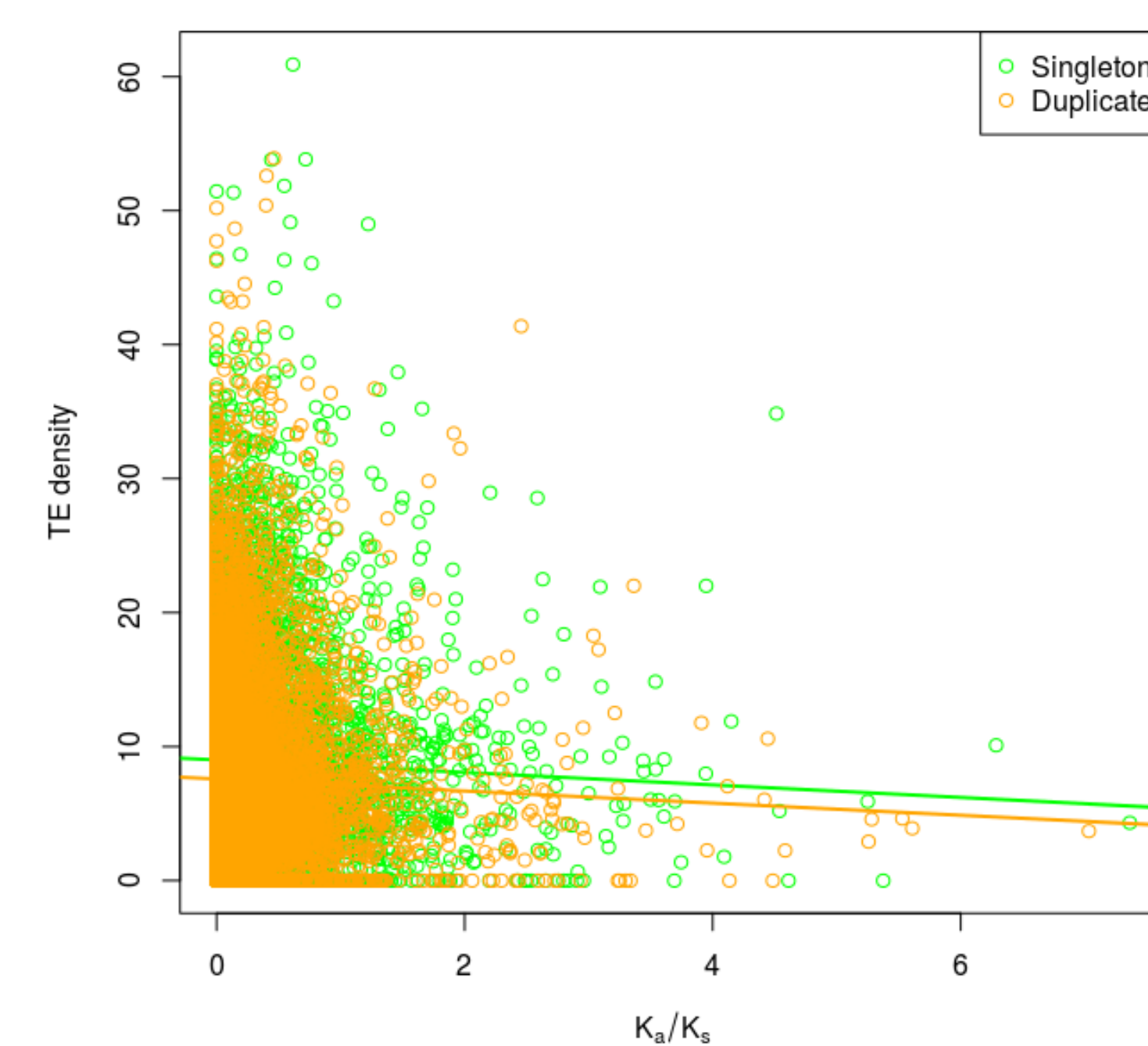


Figure 2: TE density versus  $K_a/K_s$  ratio in duplicate and singleton genes

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Gene type	1	7050	7049.6	150.9934	<2.2e-16
$K_a/K_s$	1	638	638.0	13.6646	0.0002193
Gene type: $K_a/K_s$	1	0	0.2	0.0037	0.9514887
Residuals	15131	706439	46.7		

Figure 3: Analysis of Variance Table

## CONCLUSION & PERSPECTIVES

Our results show that the context of singletons is richer in TEs than duplicates's. This can be explained by the link between gene type and selective pressure, this latter being known to be associated with the TE density. However, our results show that when adjusting for the selective pressure effect, singletons' context is still richer in TEs than duplicates, suggesting that TE context depends not only on the selective pressure but also on the duplicate status of genes. As the density of TEs is correlated with GC content (Medstrand et al., 2002), we will verify if the relationship between TE context and duplicate status is still true when taking into account the GC content in vicinity of genes. We will also explore the function of duplicate genes according to the TE context using Gene Ontology web based tool.

## REFERENCES

- Cordaux R. et Batzer MA. (2009). The impact of retrotransposons on human genome evolution. Nature.
- Medstrand, P. et al. (2002). Retroelements distributions in the human genome : Variations associated with age and proximity to genes. Genome Research.
- Ohno, S. (1970). Evolution by gene duplication. Springer.
- Sémon, M. et Wolfe, K. (2007). Consequences of genome duplication. Science Direct.
- Vinogradov, A. (2012). Large scale of human duplicate genes divergence. J Mol Evol.
- Yang, Z. et Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol.