

Supplementary Text

To validate the results obtained using the data from Sin *et al.* (2012), we performed an additional analysis on the dataset from Lesch *et al.* (2016). This dataset includes RNA-seq on sorted cells samples for pachytene spermatocytes and round spermatids, for 3 biological replicates in humans, and is available in NCBI SRA (SRP057141).

The fastq files were downloaded from NCBI. We first removed the adapters in the reads using *adapterRemoval* v1.5 (<https://github.com/MikkelSchubert/adapterremoval>). *TopHat* v2.1.1, as part of *BowTie* v2.2.8 (<http://bowtie-bio.sourceforge.net/index.shtml>), was used to map the reads on the artificial X and Y chromosomes, and *cufflinks* v2.2.1 (<http://cole-trapnell-lab.github.io/cufflinks/>) was used to process the results. We compared the FPKM levels between genes with CNV and no CNV for both pooled biological replicates and separate biological duplicates (H1, H2 and H3) in pachytene spermatocytes (PS) and round spermatids (RS).

Our results show that the X and Y-linked ampliconic genes with CNV are expressed in both PS and RS (figure S20 A and C, figure S21 A and C), except for BPY2 which is not expressed in PS (figure S21 C).

The expression of X-linked genes showing CNV is significantly higher than the expression of genes without CNV for both pooled and separate biological replicates and both cell types, as we concluded from the Sin *et al.* (2012) dataset. For the Y chromosome, the top expressed genes are TSPY, HSFY and DAZ, as observed using Sin *et al.* (2012) dataset. However, the expression of genes with CNV is not significantly higher than the expression of genes without CNV (figure S21 B and D). The differences that we observe between biological replicates indicates that expression of these genes is variable among individuals (figure S20 A-B and S21 A-B). An interesting perspective of this analysis would be to study the correlation between copy number and expression level, as copy number variations are likely to affect gene expression level.

This analysis confirms that ampliconic genes with copy number variations are indeed expressed during stages of meiosis where the sex chromosomes are transcriptionally repressed. We could also replicate our previous results: the expression of genes with CNV is significantly higher than genes with no CNV for the X-linked ampliconic genes. However, we could not replicate this result for the Y-linked ampliconic genes. Therefore, one should exert caution when interpreting the higher expression of the copy number variable Y-linked ampliconic genes as compared to the non-variable Y-linked genes, as we could not replicate it with this independent dataset.

Lesch B. J., Silber S. J., McCarrey J. R., Page D. C., 2016 Parallel evolution of male germline epigenetic poising and somatic development in animals. *Nat. Genet.* 48: 888–894.

Sin H., Ichijima Y., Koh E., 2012 Human postmeiotic sex chromatin and its impact on sex chromosome evolution. *Genome Res.*: 827–836.

Supplementary Figures

Figure S1- Copy number variation for the Y-linked ampliconic genes HSFY and XKRY

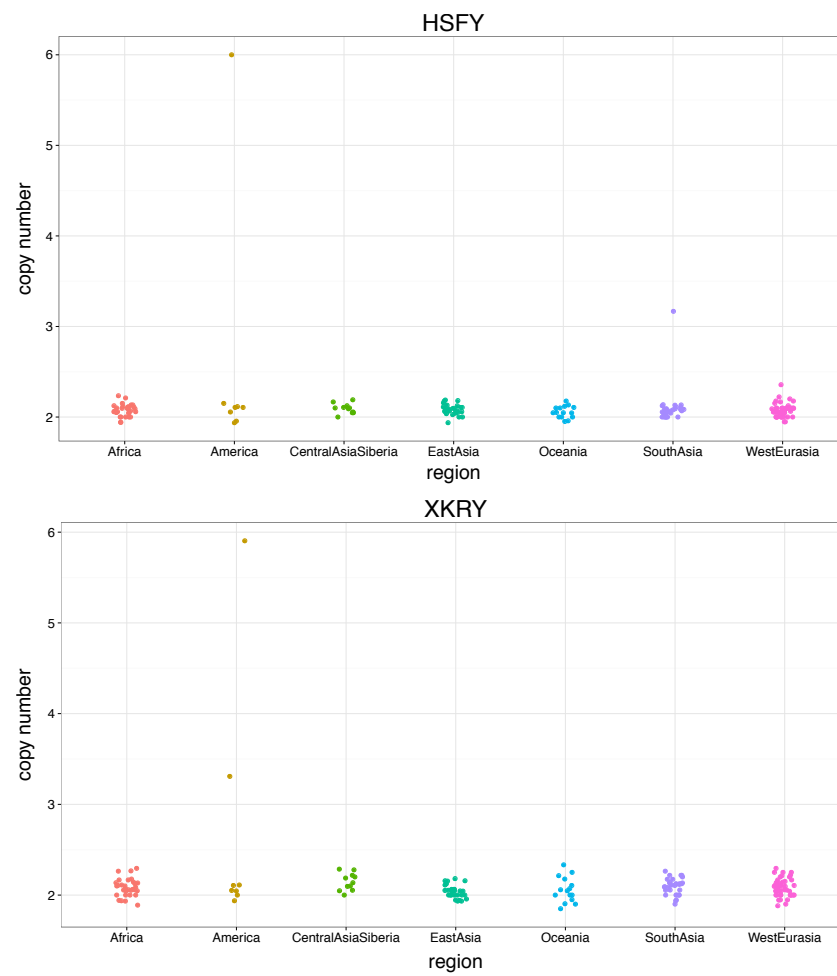
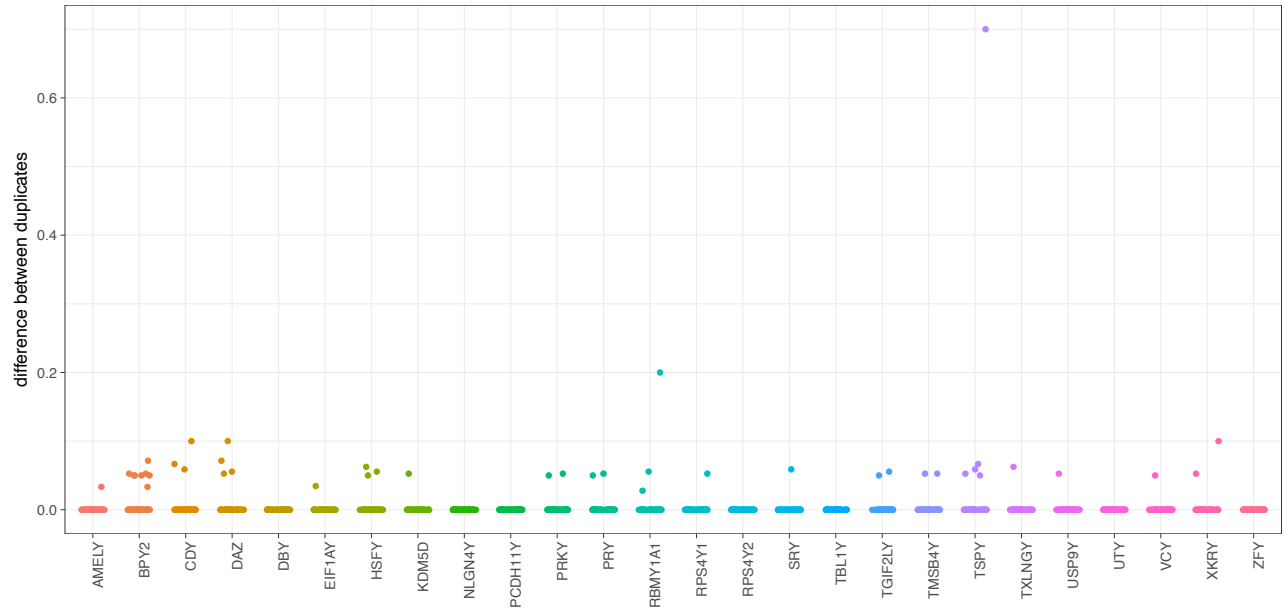


Figure S2- Difference in copy number between sequencing duplicates for A-
the Y-linked ampliconic genes and **B-** the X-linked ampliconic genes

A



B

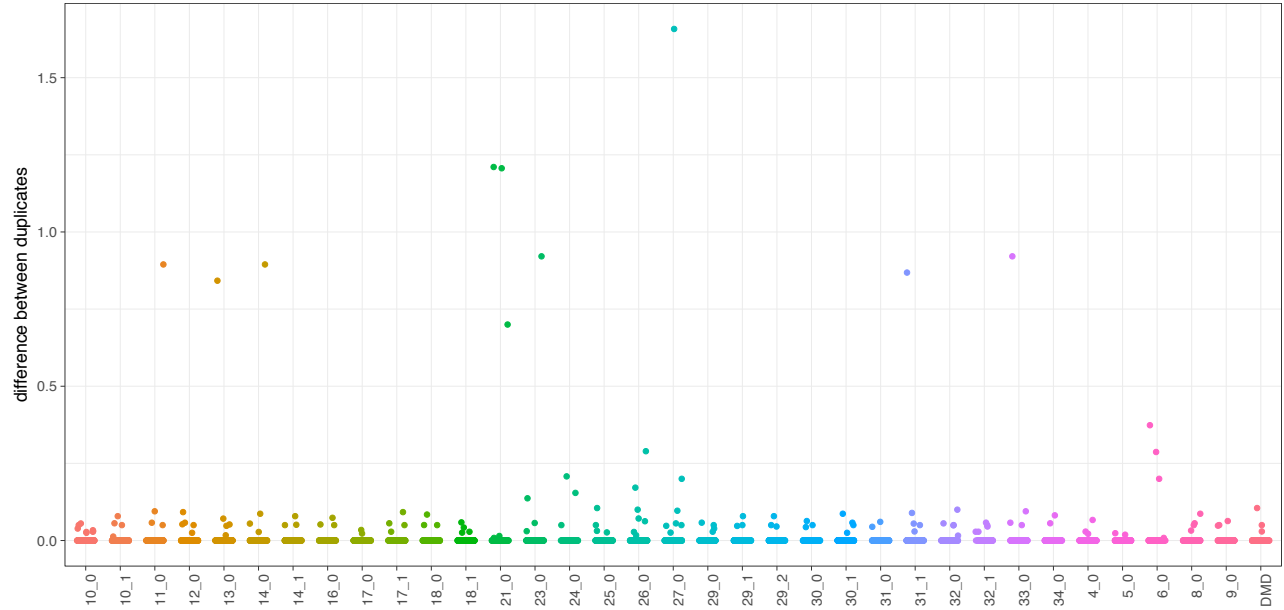


Figure S3- Length of the unit of repetition used for mapping for autosomal, X-linked and Y-linked ampliconic genes with extensive copy number variation.

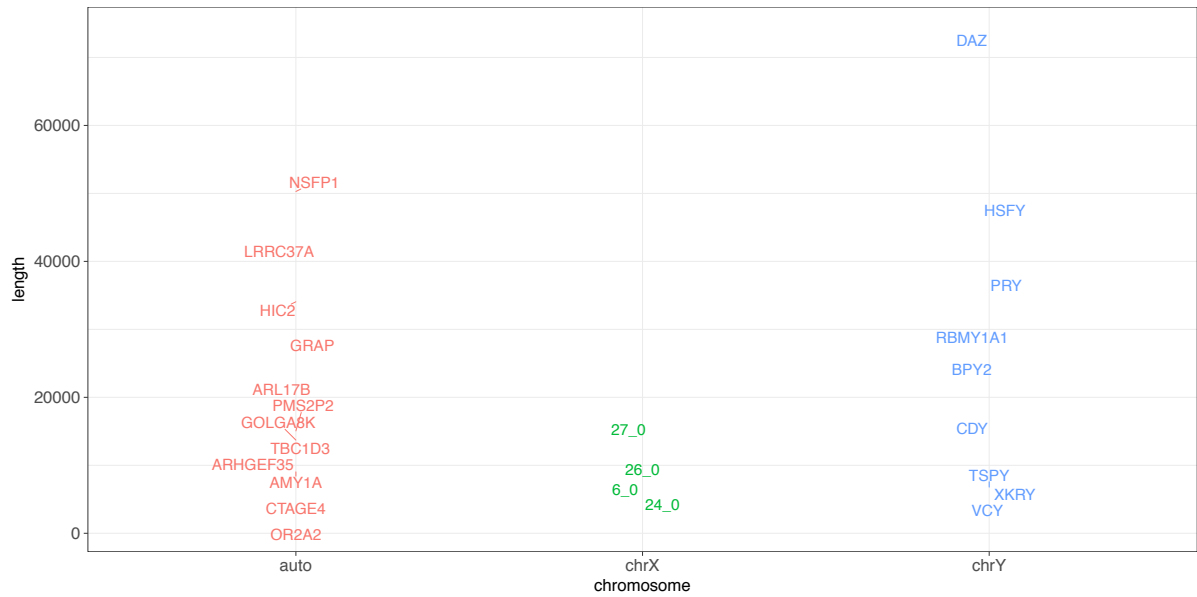


Figure S4- Copy number variation of a set of autosomal multicopy genes

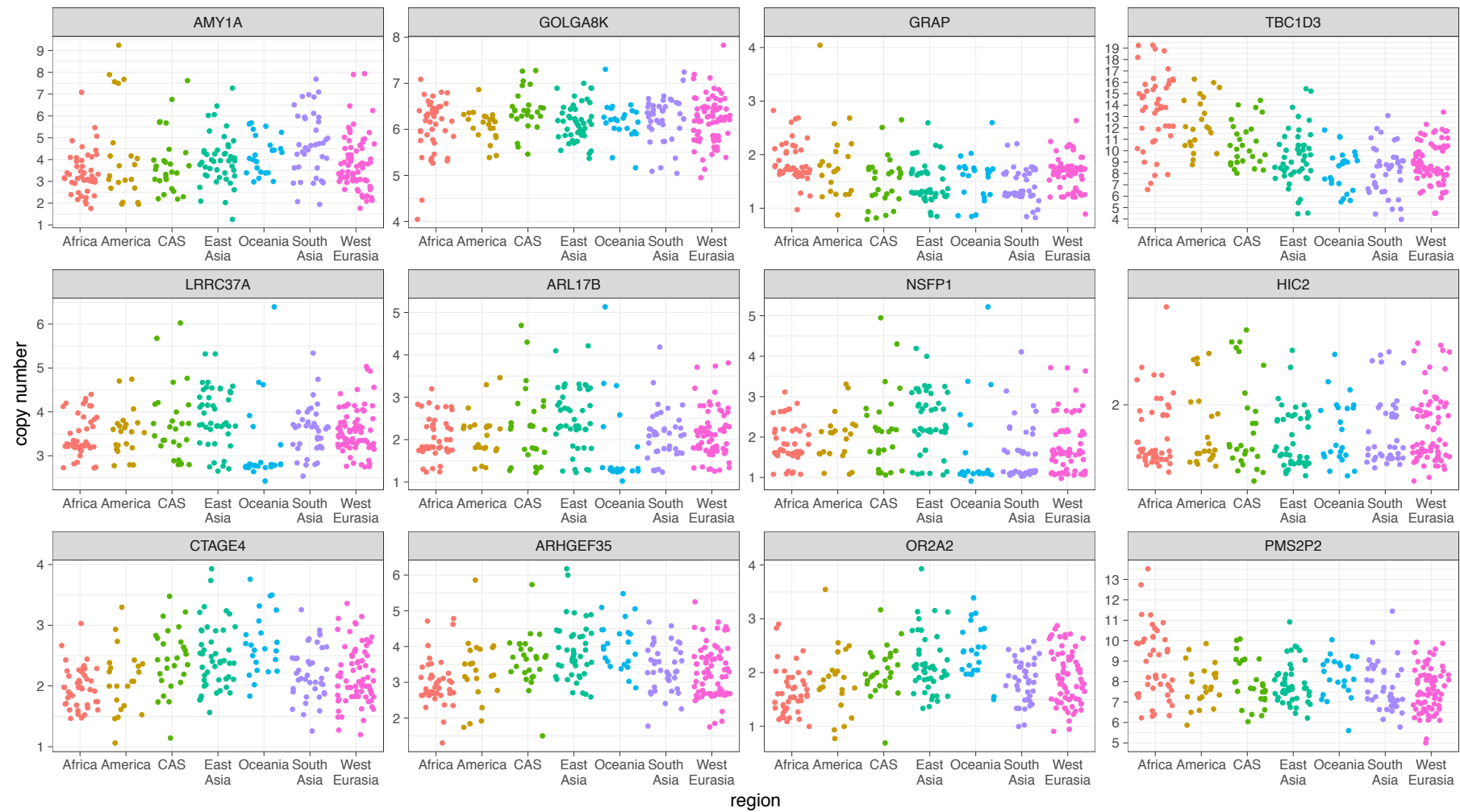


Figure S5- Maps representing each individual’s origin colored according to their copy number for the **X-linked ampliconic genes**

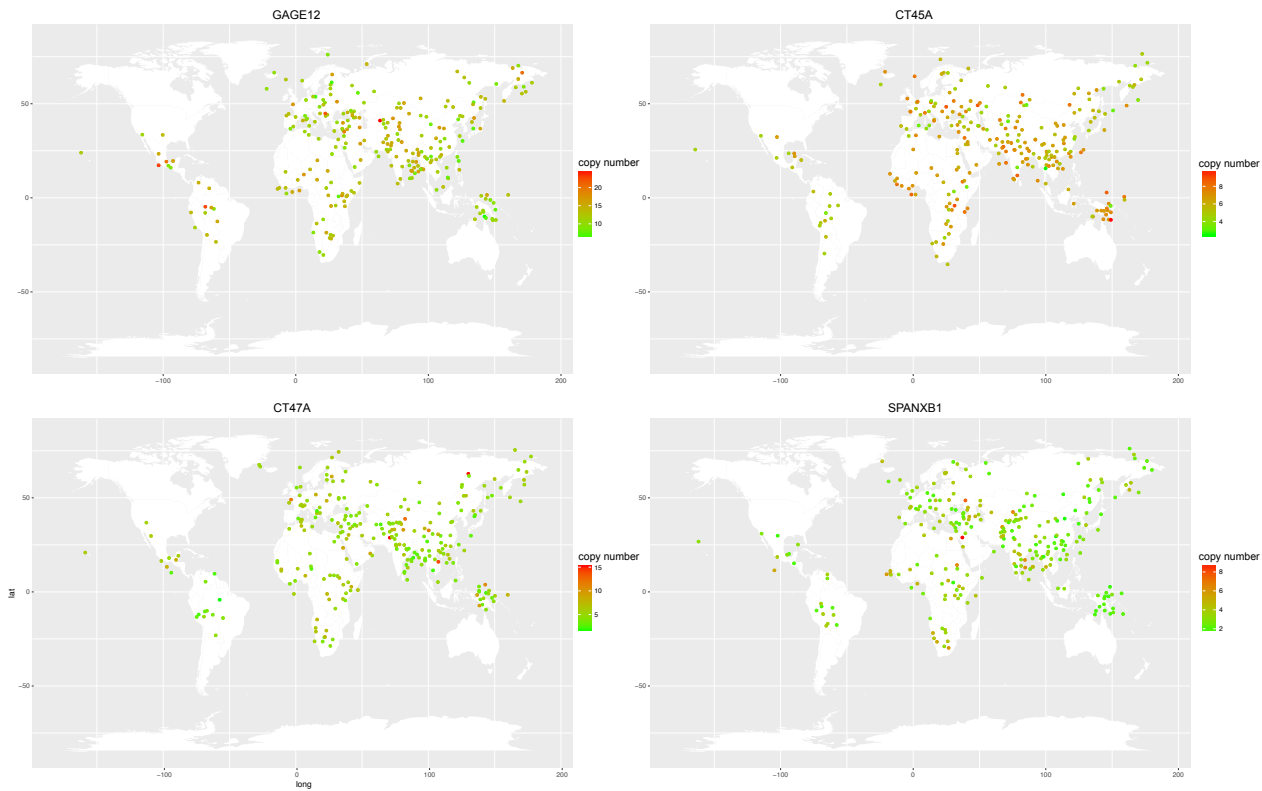


Figure S6- Maps representing each individual's origin colored according to their copy number for the **Y-linked ampliconic genes**

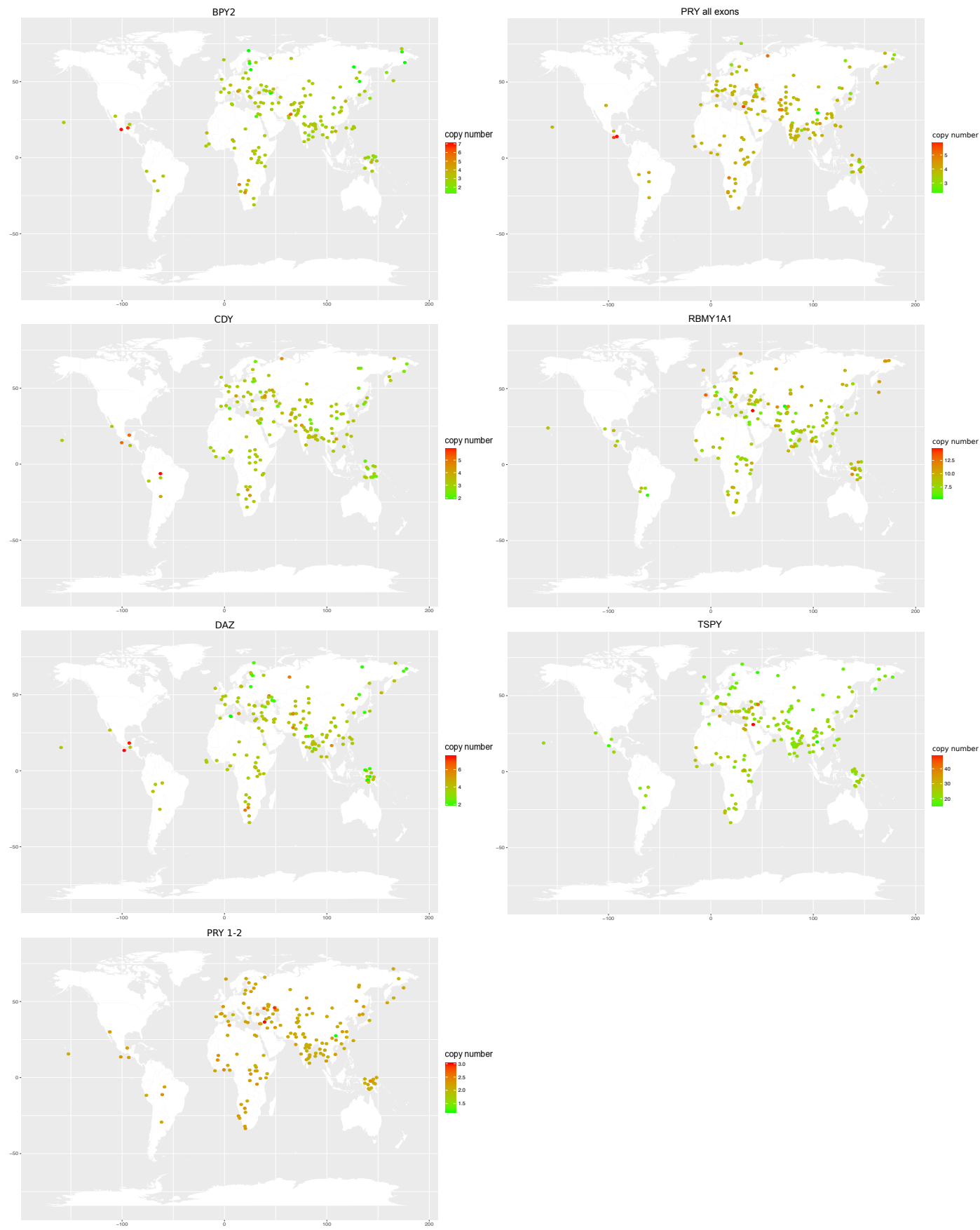


Figure S7- VST index calculated on the copy number of X- and Y-linked ampliconic genes

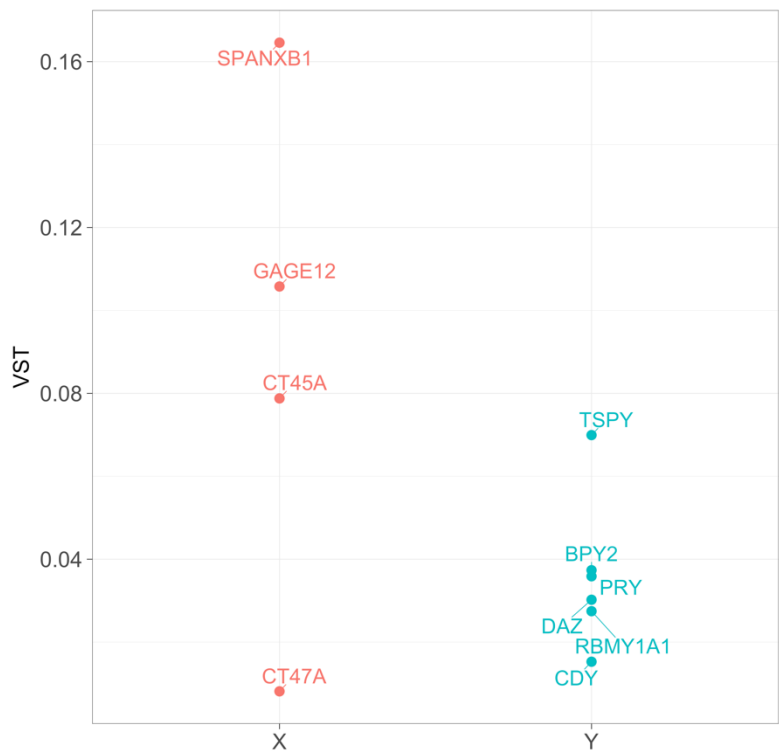


Figure S8- Copy number of Y-linked ampliconic genes. Each line is an individual. Individuals are separated by region of origin. The top panel represents the topology of the palindromes on the Y chromosome (adapted from Skov et al. 2017)

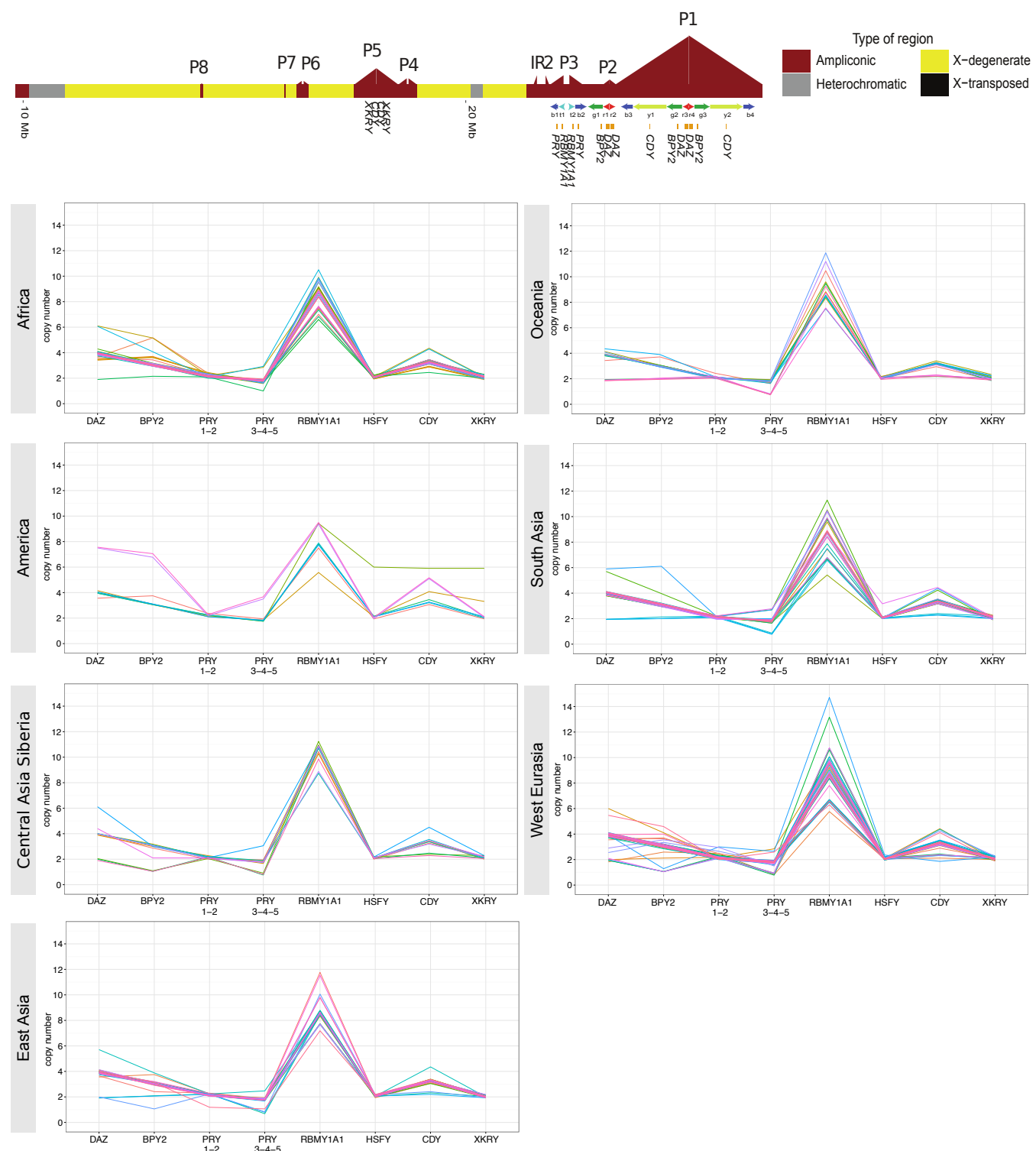


Figure S9- Illustration of the relationship between the copy number of a variant and the copy number of the gene. Individuals are indicated by a letter, from A to G, and the grey rectangles represent the copies of a specific ampliconic gene. The green box represents the derived allele of a specific SNP. The plot on the right represents the relationship between the copy number of the genes and the copy number of the derived allele of the SNP. For example, individual A has 2 copies of the gene, but no copies containing the derived allele of the variant.

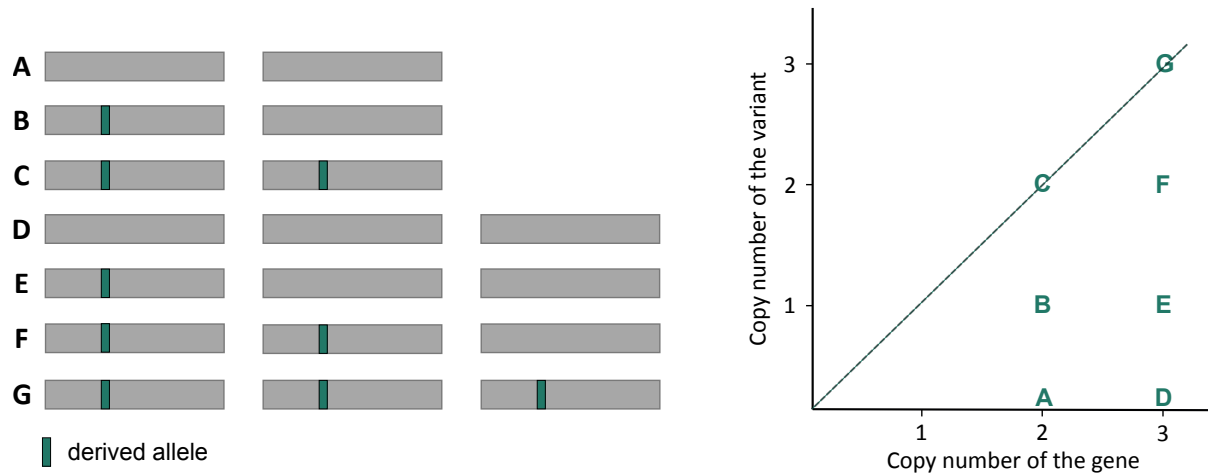


Figure S10- X-linked common NS variants. Copy number (CN) of the derived allele of each variant compared to the copy number of the ampliconic gene considered. Each point represents an individual, and is colored according to the geographical origin of the individual.

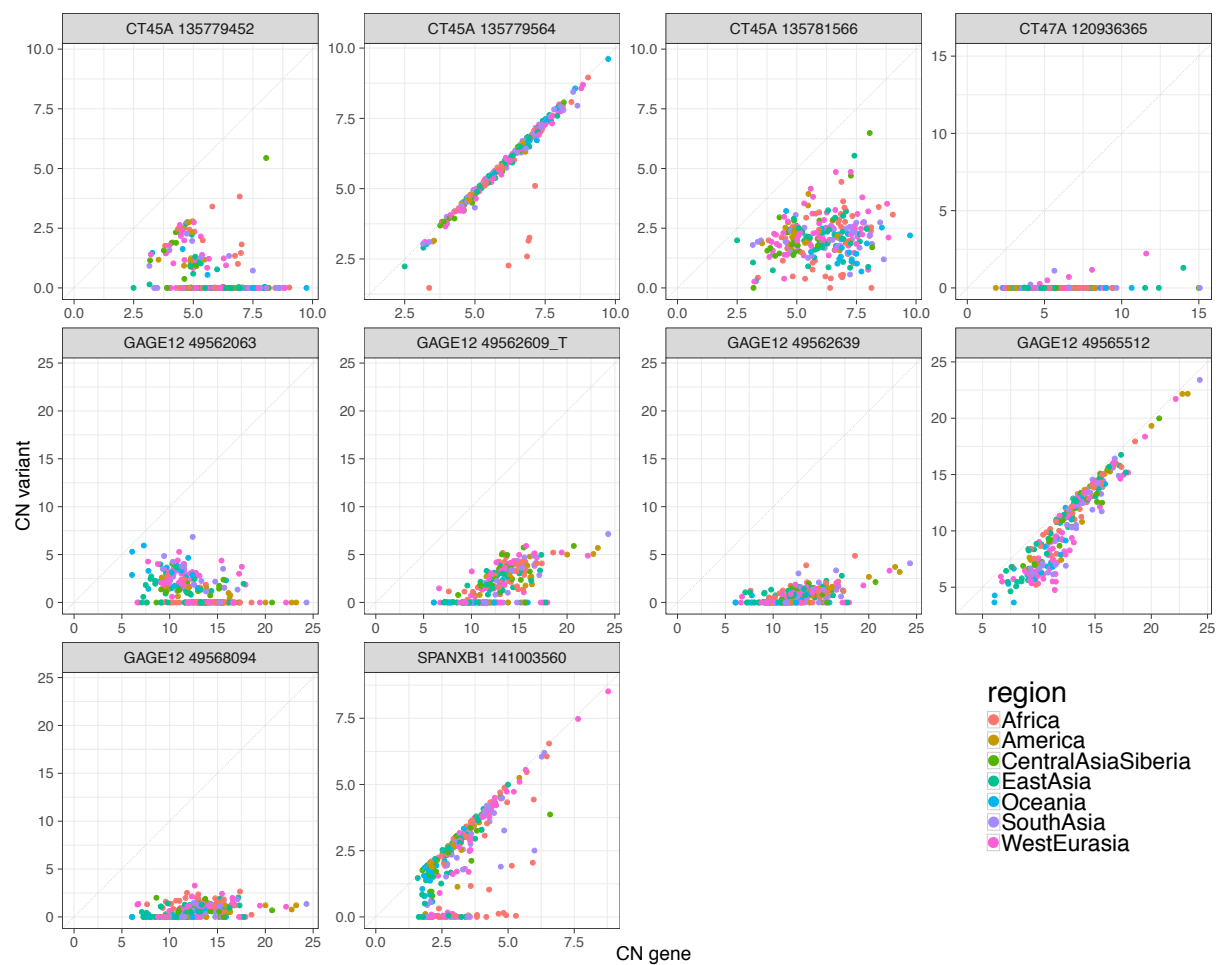


Figure S11- X-linked common S variants. Copy number (CN) of the derived allele of each variant compared to the copy number of the ampliconic gene considered. Each point represents an individual, and is colored according to the geographical origin of the individual.

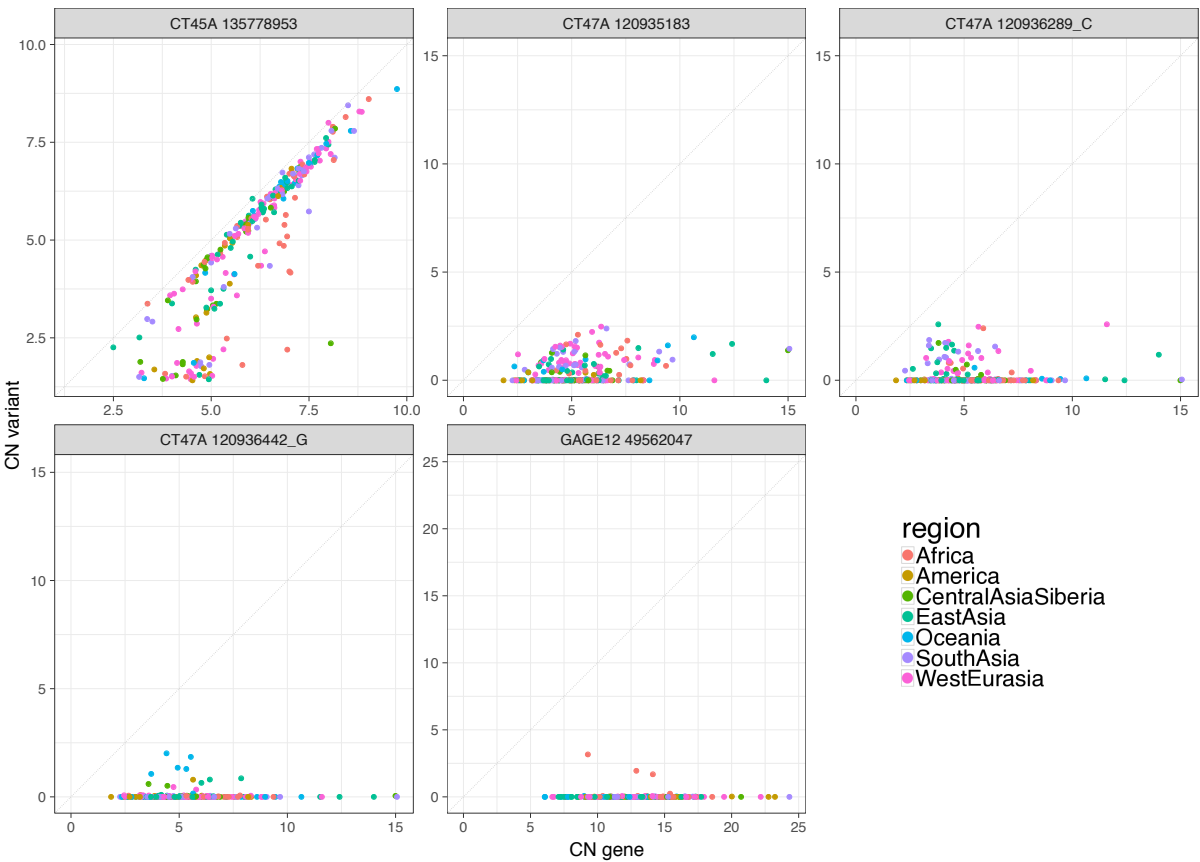


Figure S12- Y-linked common NS variants. Copy number (CN) of the derived allele of each variant compared to the copy number of the ampliconic gene considered. Each point represents an individual, and is colored according to the haplogroup of the individual. Blue boxes indicate variants for which the chimpanzee is heterozygous.

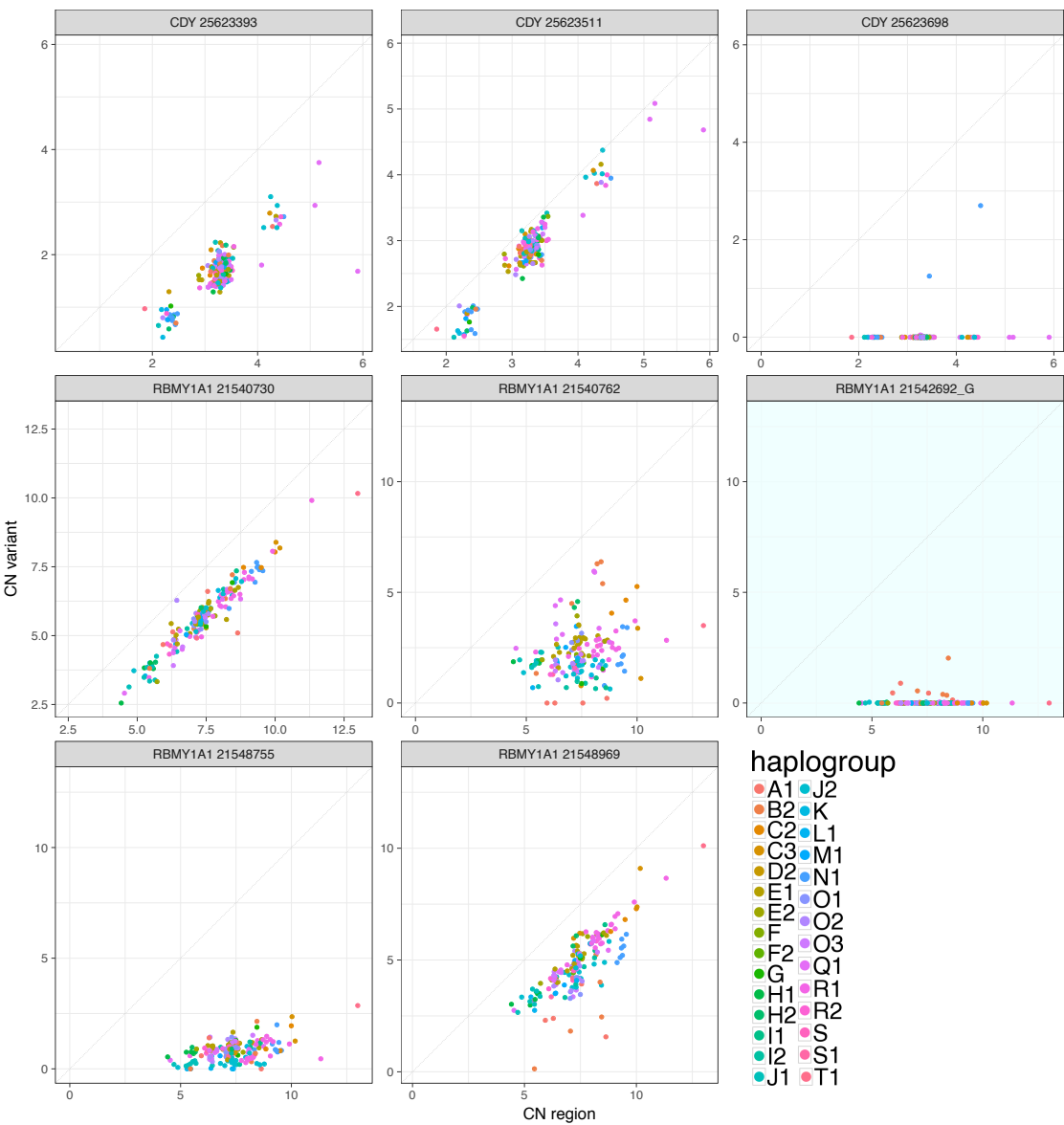


Figure S13- Y-linked common S variants. Copy number (CN) of the derived allele of each variant compared to the copy number of the ampliconic gene considered. Each point represents an individual, and is colored according to the haplogroup of the individual.

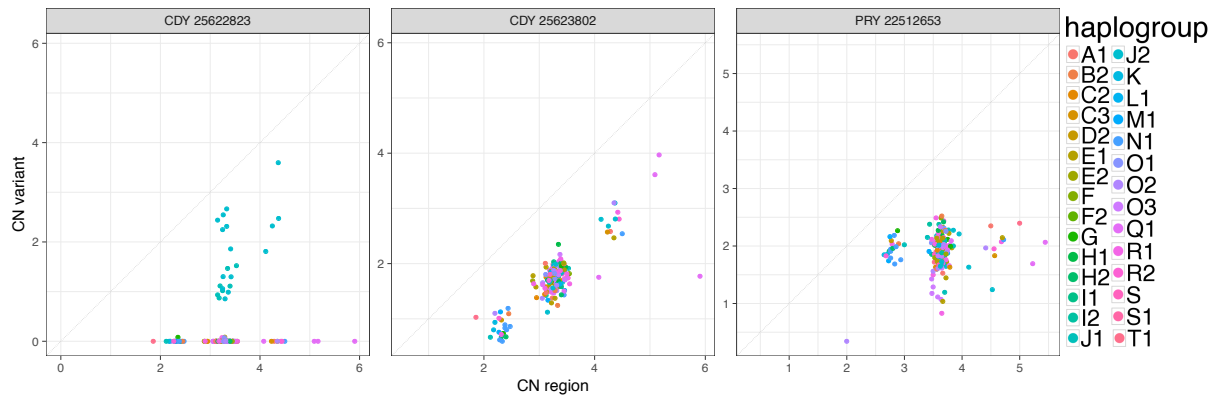


Figure S14- CDY variants. Copy number (CN) of the derived allele of each variant compared to the copy number of the ampliconic gene considered. Each point represents an individual, and is colored according to the haplogroup of the individual.

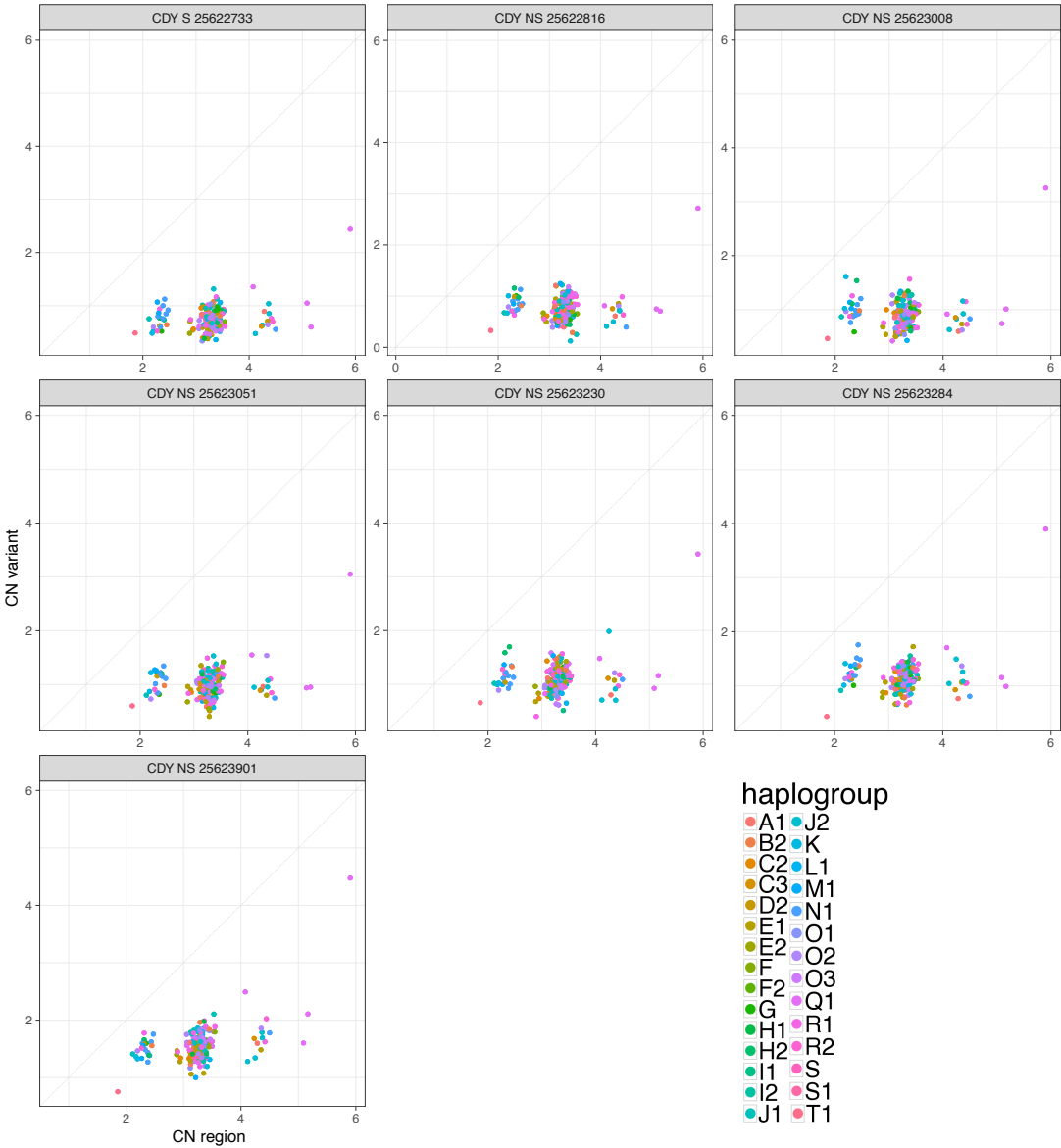
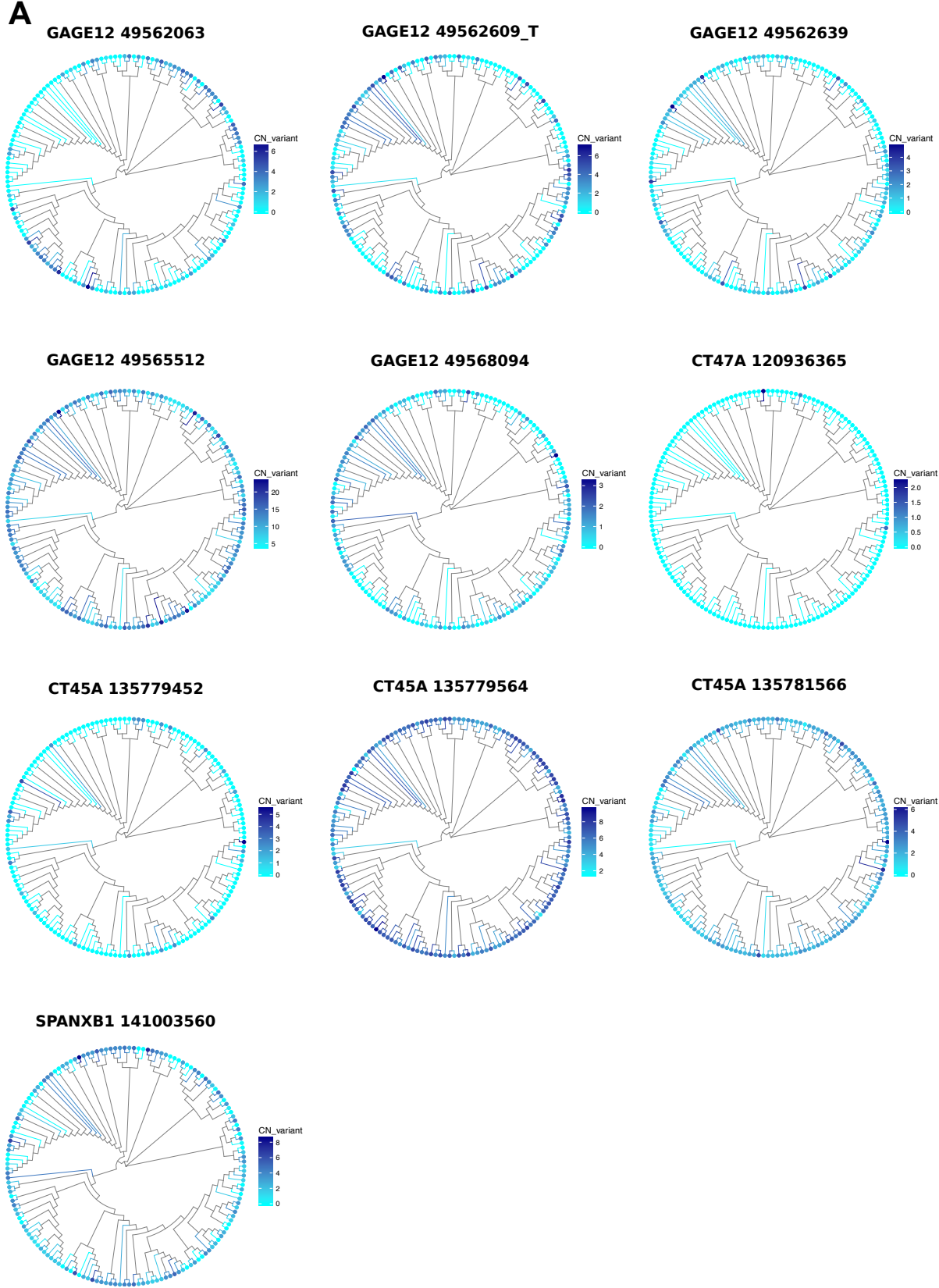
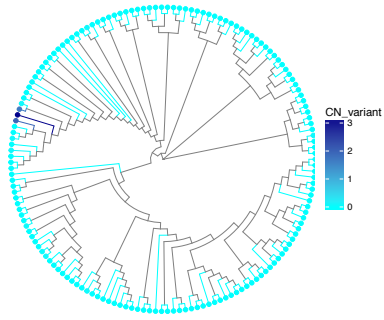
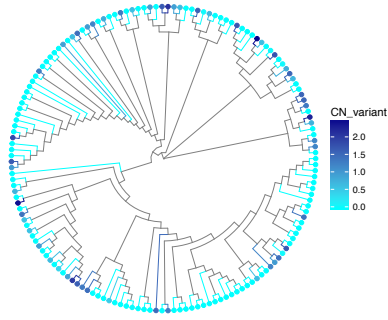
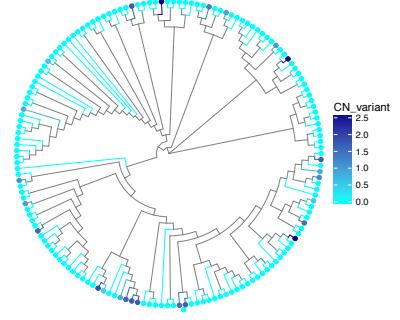
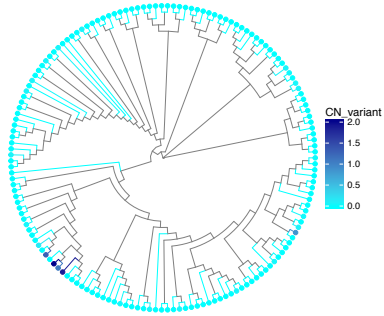
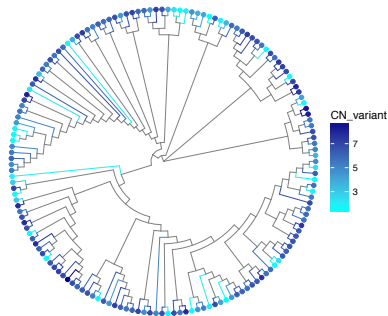


Figure S15- Tree constructed with the neighbour-joining distance between individual for the whole X chromosome and for the X-degenerate region of the Y chromosome. The tips are colored according to the copy number of the derived allele of the variant. **A-** X-linked NS variants, **B-** X-linked S variants, **C-** Y-linked NS variants, **D-** Y-linked S variants.



B**GAGE12 49562047****CT47A 120935183****CT47A 120936289_C****CT47A 120936442_G****CT45A 135778953**

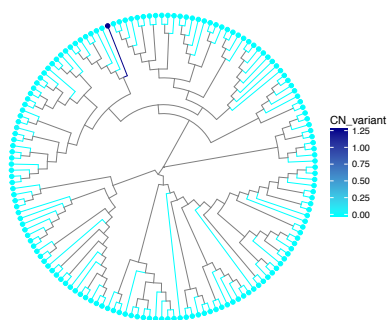
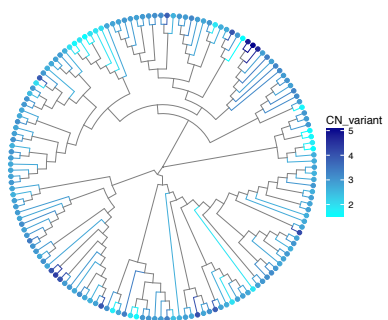
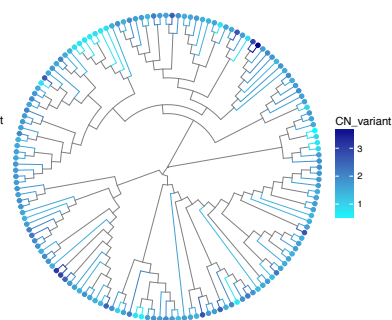
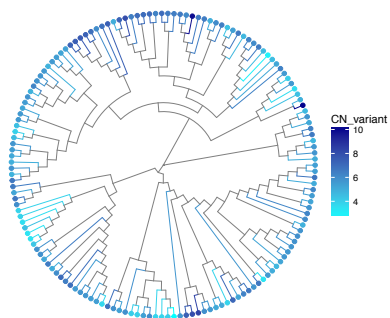
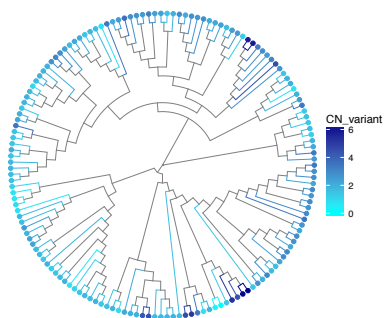
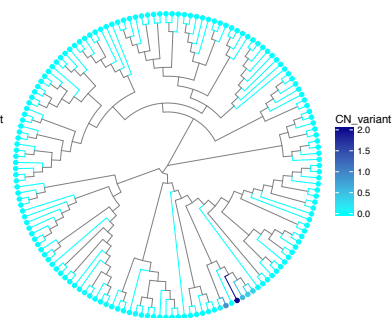
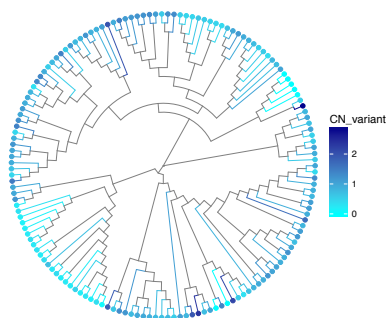
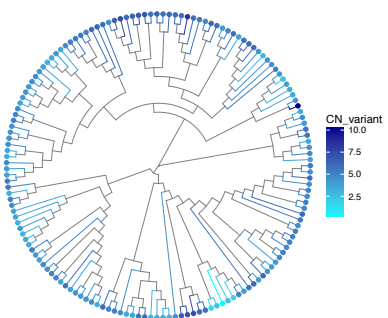
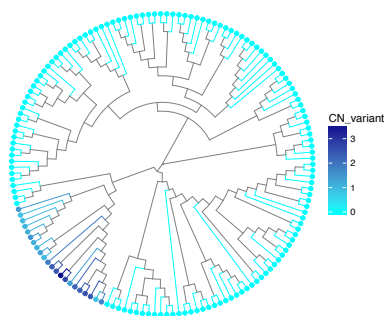
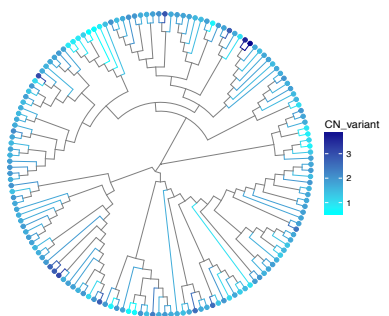
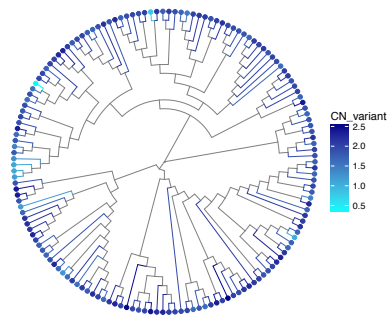
C**CDY 25623698****CDY 25623511****CDY 25623393****RBMY1A1 21540730****RBMY1A1 21540762****RBMY1A1 21542692_G****RBMY1A1 21548755****RBMY1A1 21548969****D****CDY 25622823****CDY 25623802****PRY 22512653**

Figure S16- VST index calculated for the autosomal, X-linked and Y-linked ampliconic genes copy number

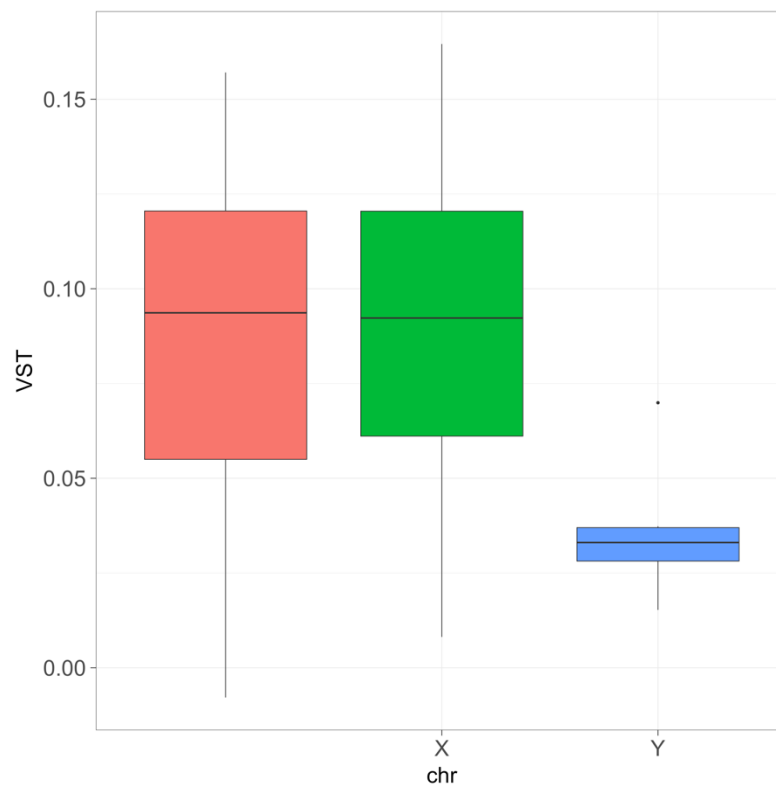


Figure S17- VST index calculated for the Non-Synonymous variant copy number of autosomal, X-linked and Y-linked ampliconic genes

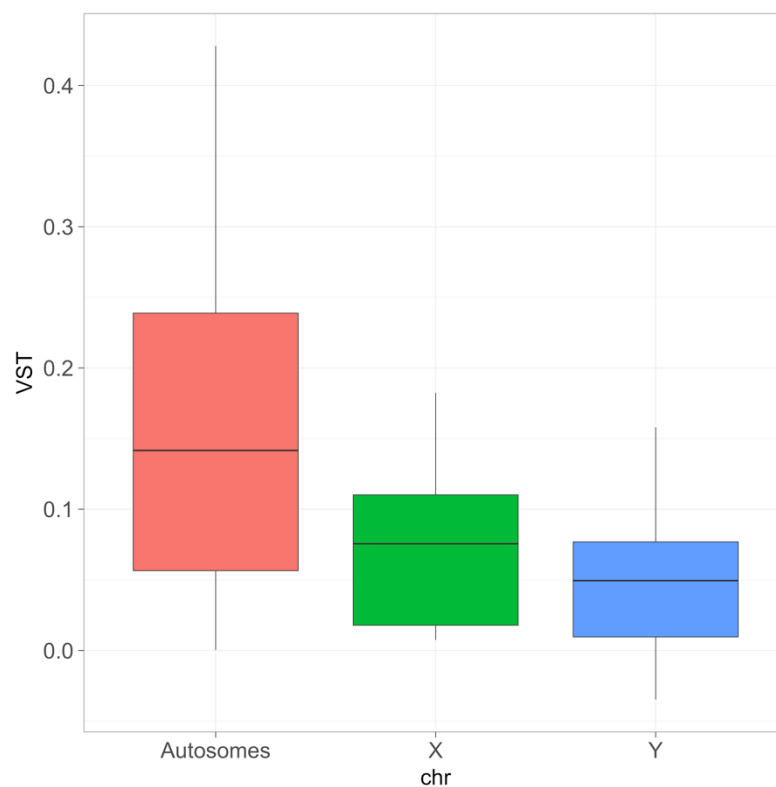


Figure S18- Median copy number as a function of variance in copy number (log scale), for the autosomal, X-linked and Y-linked ampliconic genes with extensive copy number variation.

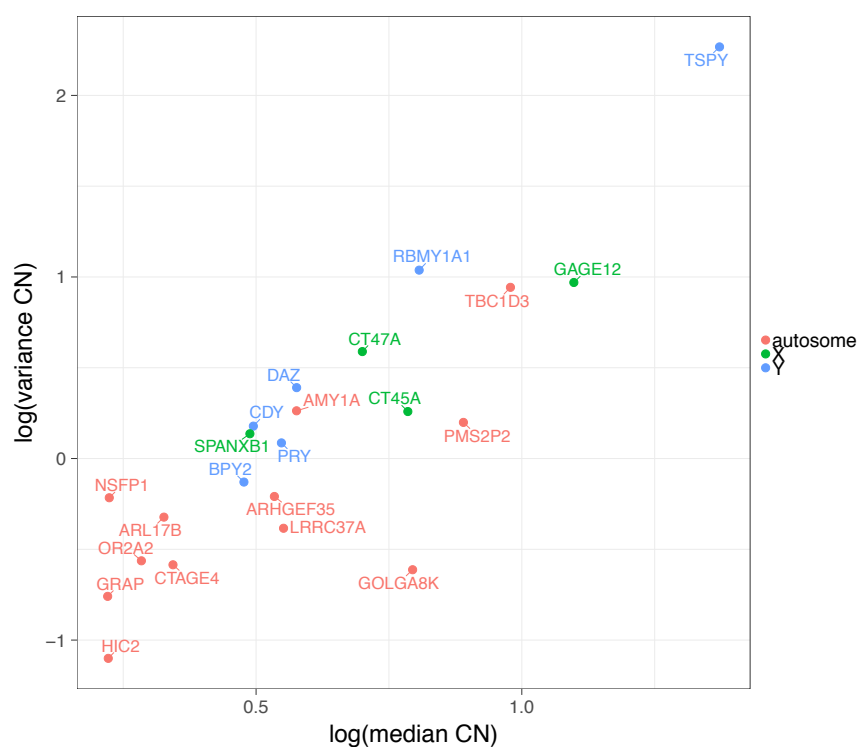


Figure S19- Median copy number as a function of variance in copy number (log scale), for NS variants located in the autosomal, X-linked and Y-linked ampliconic genes with extensive copy number variation.

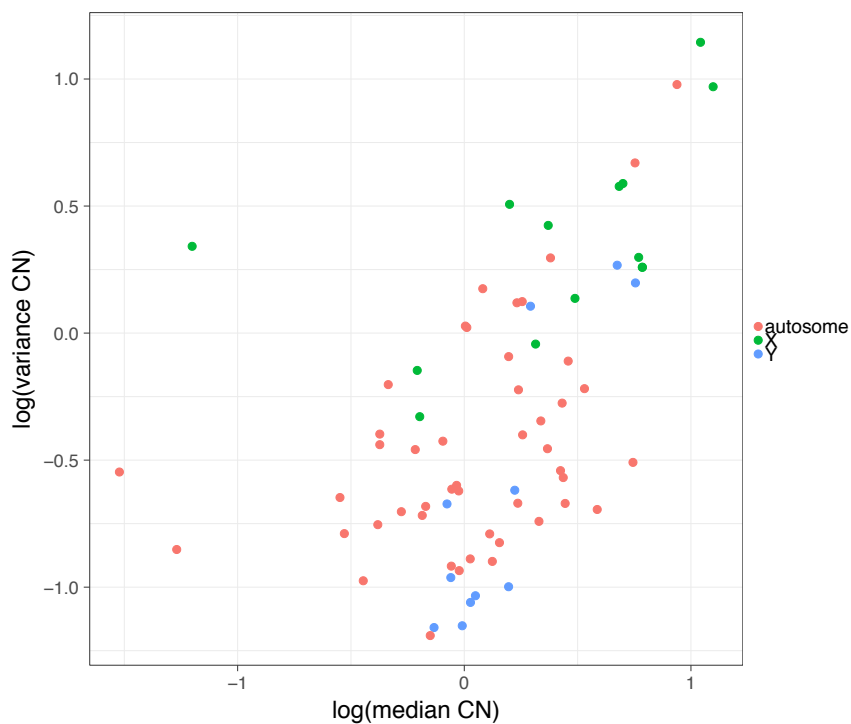


Figure S20: FPKM in pachytene spermatocytes (PS) and round spermatids (RS) for the X-linked ampliconic genes, **A-B** analyzed using the pooled biological replicates and **C-D** for each biological replicates (H1, H2 and H3); **A** and **C** only for the genes showing CNV and **B** and **D** with all the X-linked ampliconic genes.

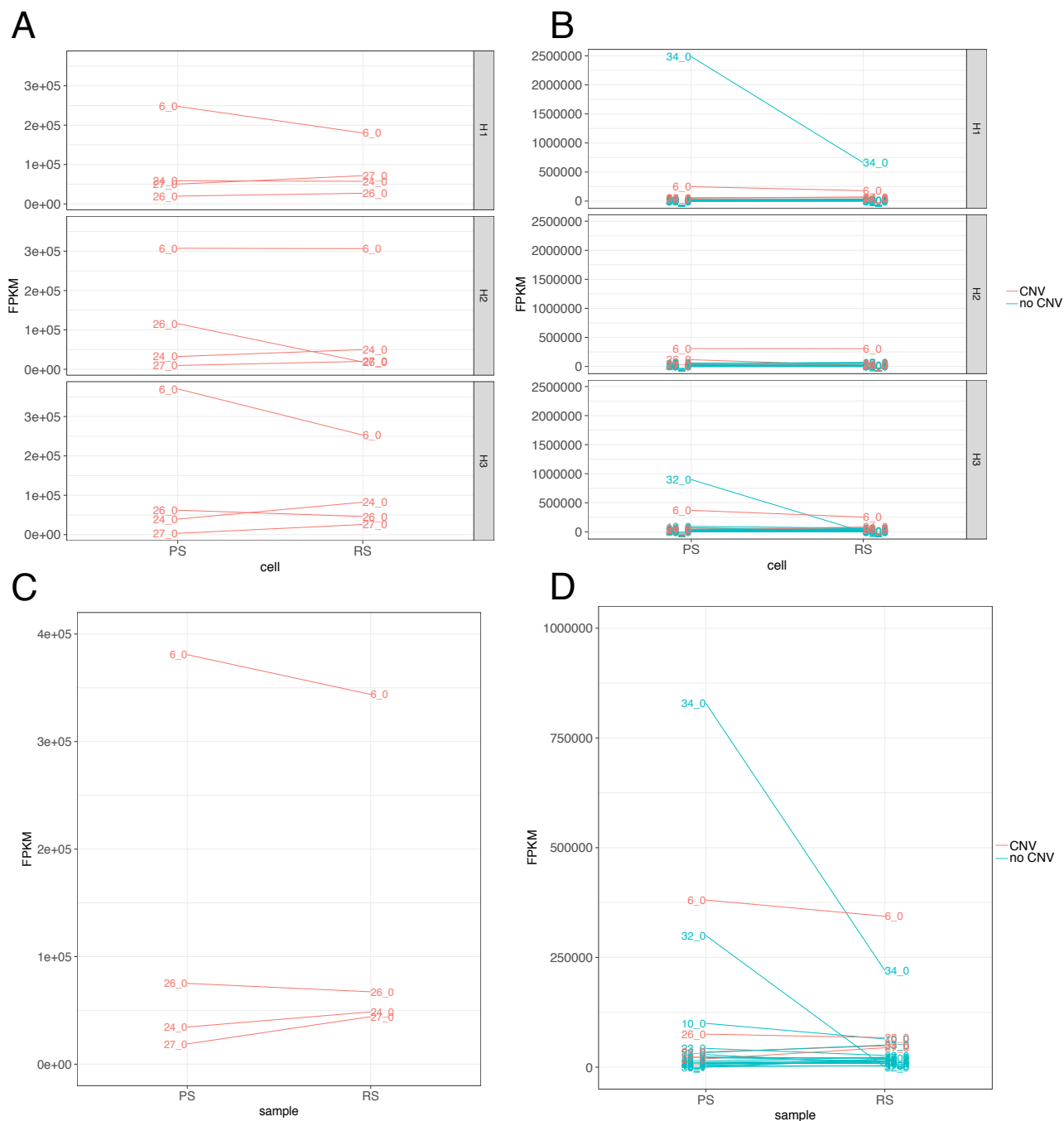
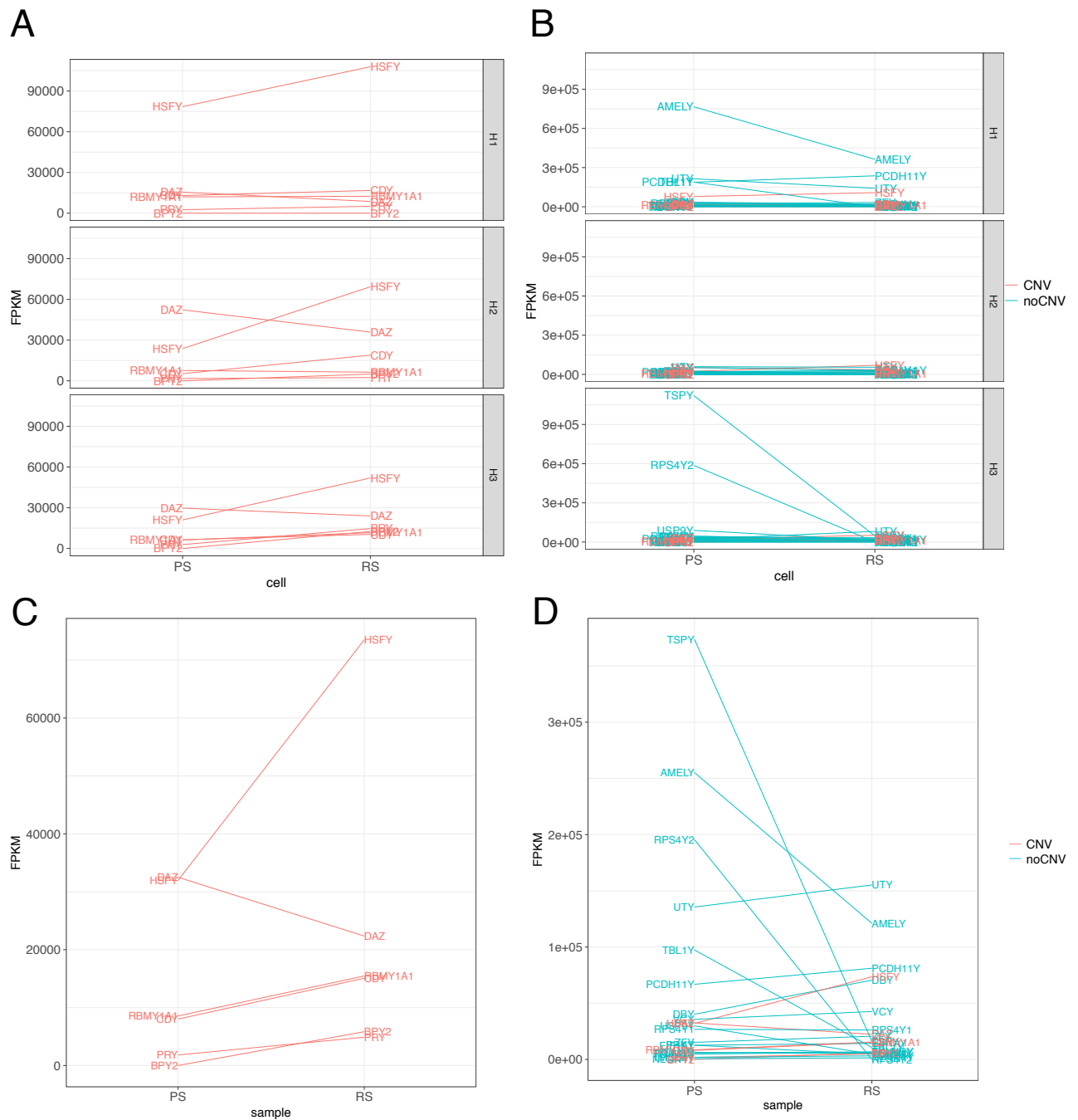


Figure S21: FPKM in pachytene spermatocytes (PS) and round spermatids (RS) for the Y-linked ampliconic genes, **A-B** analyzed using the pooled biological replicates and **C-D** for each biological replicates (H1, H2 and H3); **A** and **C** only for the genes showing CNV and **B** and **D** with all the Y-linked genes included in our study.



Supplementary Tables

Table S1- Summary of the copy number variation for the X-linked ampliconic regions- See Excel file

Table S2- Summary of the copy number variation for the Y-linked ampliconic genes- See Excel file

Table S3- Set of autosomal ampliconic regions selected. From Sudmant *et al.* (2015), Dennis *et al.* (2016) and Warbuton *et al.* (2004)

from	type	chromosome	gene
Sudmant <i>et al.</i> (2015)	CNV	chr6	CLPS
		chr16	HP, HPR
		chr12	DPY19L2
		chr1	LRRIQ3
		chr17	CRHR1
		chr2	ANKRD36
		chr1	NBPF
Dennis <i>et al.</i> (2016)	segmental duplication	chr7	PMS2L5
		chr7	SPDYE1
		chr7	DPY19L2
		chr9	SPATA31A1
		chr9	ZNF790
		chr11/chr2	TRIM51
		chr15	GOLGA
		chr16	NPIP
		chr17	LRRC37A
		chr17	TBC1D3
Warburton <i>et al.</i> (2004)	large inverted repeats	chr1	AMY1A, AMY2A
		chr2	TRIM-43
		chr7	BC036215
		chr7	BC043153
		chr7	AF327904
		chr11	RNF18
		chr15	AK09040
		chr17	GRAP
		chr22	HIC2, AL133030

Table S4- Summary of copy number variation for the autosomal ampliconic genes selected after performing the dotplots- See Excel file