Sensory receptors

Ref. 2,
The authors should provide some direct evidence that the highlighted peak corresponds to sensory genes.
For example, they could show that the peak disappears when sensory genes are excluded

El pico en la distribución lo observé hace muchos años al trabajar sobre la lognormal de la distribución de longitudes de genes. El objeto de trabajo eran los pseudogenes. Me quedó claro que, en humanos, muchos pseudogenes (unitary psgs) que fueron "olfatory receptors" estaban en el pico. Las anotaciones de genes entonces estaban peor que ahora ya que hoy en día esos psgs ya no se confunden con genes que codifican proteína.

No podemos filtrar por la descripción del gen o el nombre del gen, para buscar los receptores (sensory perception) buscaré en la ontología de genes (GO) los distinto términos asociados. Finalmente, para la especie que es objeto del estudio, representaré la distribución con/sin los "sensory receptor genes".

Protein coding genes

Procedure:

1. I used AmiGo to seach all the GO (Gene Ontology terms) associated with: 1.1 "sensory perception" (biological function). There are 77 GO terms. 1.2. "receptors". There are 2325 from different GO sources

2. Biomart has been used, the version corresponding ot ensembl 98 to find all the genes annotated as sensory perception genes. For that, all the genes of the species will be filtered to those with "sensory perception" GO terms (1110 diff genes).

3. Biomart has been used, the version corresponding ot ensembl 98 to find all the genes annotated as receptor genes. For that, all the genes of the species will be filtered to those with "receptor" GO terms (5913 diff genes).

4. Use the 998 "sensory perception receptors" to analyze the peak in the distribution.

Results

All	SR-genes	All but SR-genes
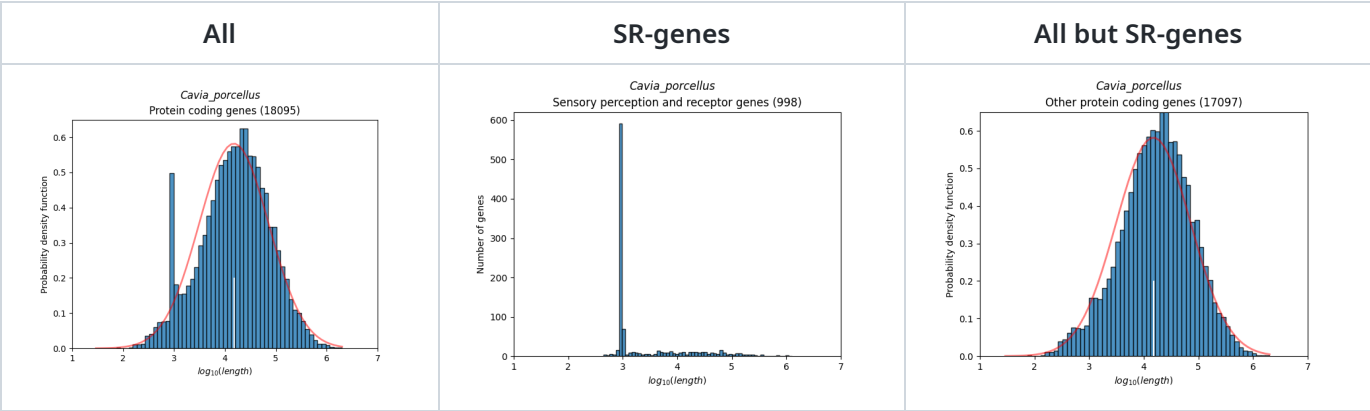
Proteins

Procedure:

1. Using Biomart, the version corresponding ot ensembl 98, and the set of 998 protein coding genes annotated (see the previous section) as sensory perception receptors to find their corresponding uniprot ids. 1123 trembl and swissprot ids are obtained.

2. Finally, use the ids them to analyze the peak in the distribution. Its corresponds again (very good) to 998 genes

Results

All	SR-prots	All but SR-prots