

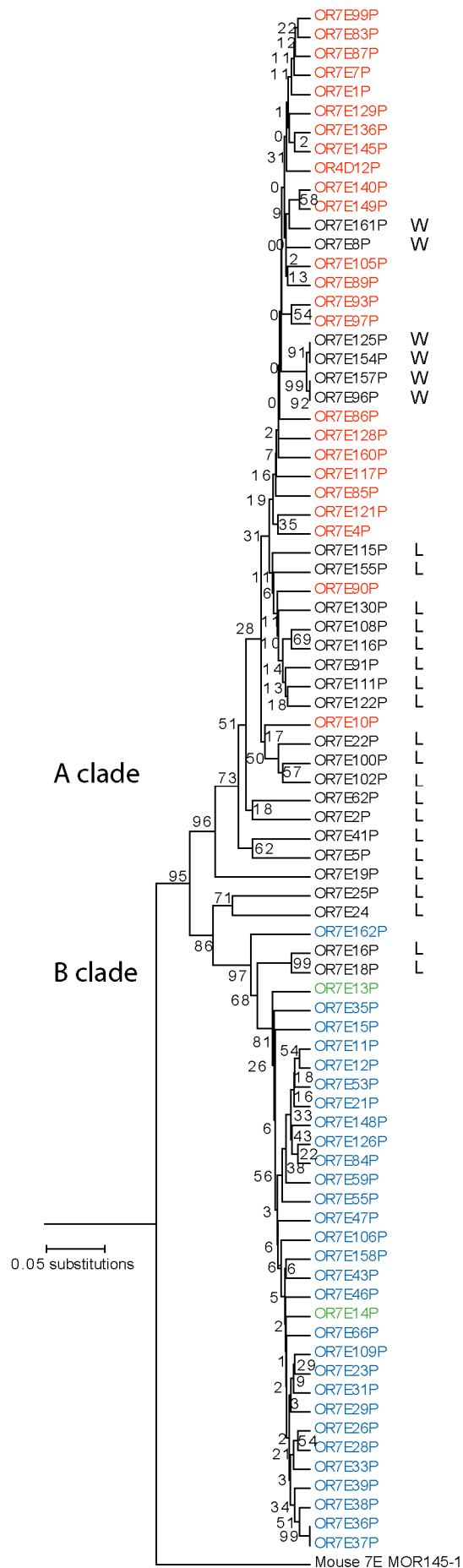


Figure Legend

A UPGMA tree of 81 human 7E nucleotide sequences from UCSC Assembly build 35, hg 17, and one mouse 7E ortholog. Green names denote pseudogenes with a common frameshift in TM3 leading to a premature stop codon. Red denotes pseudogenes with a substitution resulting in a shared stop codon in TM6 (TGA). Blue identifies pseudogenes with both the TM3 frameshift and TM6 stop codon. Black indicates 7E sequences with neither of these mutations. The letters "L" and "W" differentiate between TTA (leucine) and TGG (tryptophan) at the TM6 codon, respectively. Six 7E sequences spanning less than 70% of the alignment are omitted from this analysis. We include OR4D12, which, despite a HORDE name indicating membership in the 4D family, clusters within the 7E family in our analyses. The tree was constructed using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) (Sneath and Sokal, 1973) as implemented by MEGA3 (Kumar et al., 2004) (www.megasoftware.net). Values at nodes indicate bootstrap values as per cent of 1000 replications. Size bar, 0.05 substitutions, estimated with constant rate for all lineages. The nomenclature is based on The Human Olfactory Receptor Data Exploratorium (HORDE) version 41 (<http://bioportal.weizmann.ac.il/HORDE/>); coordinates of human sequences in May 2004 assembly (hg17) are supplied in Supplementary Table A.

Kumar, S., Tamura, K., and Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 5, 150-163.

Sneath, P. H. A., and Sokal, R. R. (1973). *Numerical Taxonomy* (San Francisco: W.H.