Supplemental Data

Extensive Copy-Number Variation of the Human Olfactory Receptor Gene Family

Janet M. Young, RaeLynn M. Endicott, Sean S. Parghi, Megan Walker, Jeffrey M. Kidd, and Barbara J. Trask

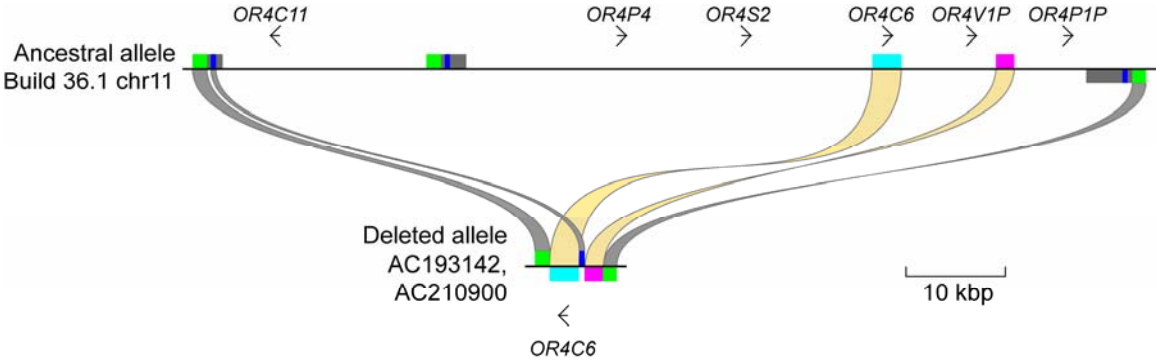

Figure S1. Alternative Structural Alleles of a CNV Region of Chromosome 11 Containing Several ORs, Including *OR4C11*

A complex combination of at least four deletions with inversion(s) separates the two structures. The ancestral allele (prevalence 65%) is represented by sequence from the reference human genome assembly, NCBI Build 36.1, (the Venter assembly provides sequence across most of this region, confirming authenticity of this structure) and the minor allele by fosmid sequence AC193142 (fosmid AC210900, derived from a different individual, has the same structure, confirming AC193142's rearrangement). The four deletions have sizes 372 bp, 66.1 kb (containing *OR4C11*, *OR4P4* and *OR4S2*), 9.5 kb (containing *OR4V1P*) and 11.9 kb (containing *OR4P1P*).

MLPA results for *OR4C11* confirm that 6 individuals have a homozygous deletion of this gene, 16 are heterozygously deleted, and 21 appear to have the "normal" two copies. PCR analysis of other genes in the region shows that the 6 individuals who are homozygously deleted for *OR4C11* are also missing *OR4P4*, *OR4S2* and *OR4P1P*, and that all individuals with at least one copy of *OR4C11* also have at least one copy of those three genes. PCR also shows that *OR4C6* is present in all individuals in our panel. Although we have not formally proven that all deleted individuals share the same structure as the two fosmids, our results strongly support the idea that the genomic structure represented by these fosmids is indeed responsible for the *OR4C11* deletions observed by MLPA.

Arrowheads represent OR genes, and colored rectangles represent blocks of continuous sequence shared by the two alleles. Gray/blue/green rectangles represent L1 sequences present in three highly similar copies in the region of the reference assembly shown (as well as additional copies nearby). Rectangles above the horizontal line represent forward strand similarity, and rectangles below the horizontal line represent inverted copies.

Curved, shaded lines connect identical sequences in the two allelic structures, with yellow shading denoting inverted sequence. Pairing might have initiated between the two L1 copies flanking the region, perhaps creating some kind of loop structure, but deletion/inversion breakpoints all suggest non-homologous end-joining was responsible for creating the final structure.

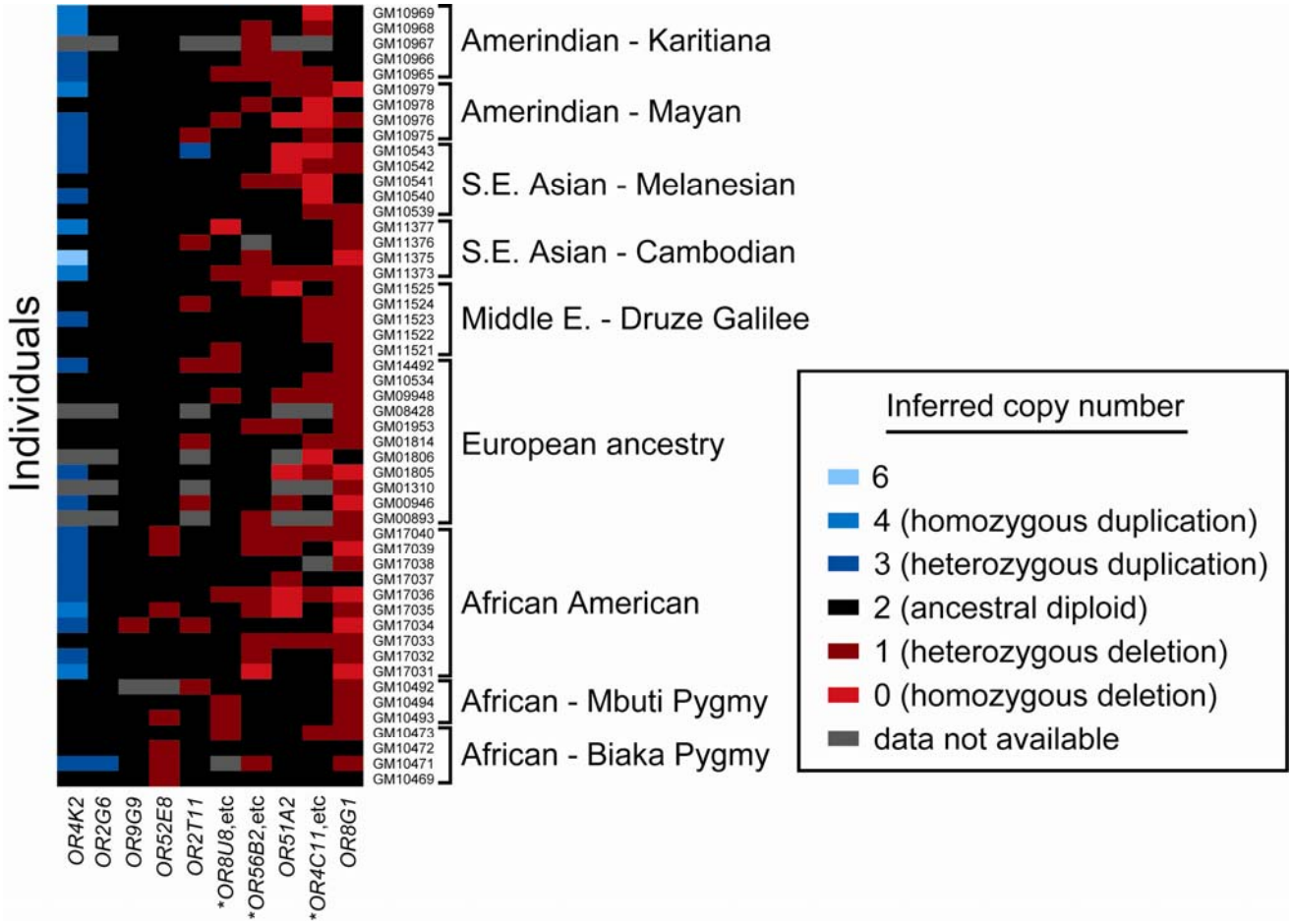

Figure S2. Number of Genomic Copies of 19 OR Genes Examined in a Panel of 51 Individuals

This figure differs from Figure 2 in that it shows only copy-number variation and ignores the effect of single nucleotide polymorphisms that would disrupt OR function. Figure 2 includes SNP data. Each row represents one of the human individuals tested as part of our diversity panel. Each column of the grid summarizes genotype data for an OR gene, or in three cases (*) for groups of OR genes (see below). Genes are ordered according to number of copies gained or lost, averaged over the individuals surveyed.

*: "*OR8U8*,etc": a deletion CNV destroys function of *OR8U8* and *OR8U9*, while simultaneously creating a novel hybrid gene, *OR8U1*; "*OR56B2*, etc": a deletion removes all of *OR56B2* and *OR52N5* as well as half of *OR52N1*; and "*OR4C11*, etc": a complex set of deletions removes *OR4C11*, *OR4P4*, *OR4S2*, *OR4V1P* and *OR4P1P* (Figure S1).