Evolutionary dynamics of olfactory receptor genes in Drosophila species

Masafumi Nozawa, and Masatoshi Nei

PNAS published online Apr 16, 2007;
doi:10.1073/pnas.0702133104

This information is current as of April 2007.

Supplementary Material
Supplementary material can be found at:
www.pnas.org/cgi/content/full/0702133104/DC1
This article has been cited by other articles:
www.pnas.org#otherarticles

E-mail Alerts
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

Rights & Permissions
To reproduce this article in part (figures, tables) or in entirety, see:
www.pnas.org/misc/rightperm.shtml

Reprints
To order reprints, see:
www.pnas.org/misc/reprints.shtml

Notes:
Olfactory receptor (OR) genes are of vital importance for animals to find food, identify mates, and avoid dangers. In mammals, the number of OR genes is large and varies extensively among different orders, whereas, in insects, the extent of interspecific variation appears to be small, although only a few species have been studied. To understand the evolutionary changes of OR genes, we identified all OR genes from 12 Drosophila species, of which the evolutionary time is roughly equivalent to that of eutherian mammals. The results showed that all species examined have similar numbers (~60) of functional OR genes. Phylogenetic analysis indicated that the ancestral species also had similar numbers of genes, but there were frequent gains and losses of genes that occurred in each evolutionary lineage. It appears that tandem duplication and random inactivation of duplicate genes are the major factors of gene number change. However, chromosomal rearrangements have contributed to the establishment of genome-wide distribution of OR genes. These results suggest that the repertoire of OR genes in Drosophila has been quite stable compared with the mammalian genes. The difference in evolutionary pattern between Drosophila and mammals can be explained partly by the differences of gene expression mechanisms and partly by the environmental and behavioral differences.

Results

Numbers of OR Genes in Drosophila Species. Our homology search (see Materials and Methods) detected 711 functional, 67 nonfunctional, and 34 partial OR genes in the genomes of 12 Drosophila species. All species examined have similar numbers of functional OR genes and much smaller numbers of pseudogenes (Table 1). Although simulans, sechellia, persimilis, and viridis show somewhat smaller numbers of functional genes, they have larger numbers of pseudogenes and partial genes. Here a partial gene refers to a gene with an open reading frame (ORF) truncated at the end of a genomic contig studied. This gene may therefore become a functional gene when the entire genomic sequence is assembled. Some pseudogenes might also become functional genes later because the nonsense or frameshift mutations identified could be caused by sequencing errors. For these reasons, our estimates of functional genes are likely to be minimums except in melanogaster, where the genome sequence is well established. In particular, the Hawaiian fruit fly grimshawi, which has many pseudogenes and partial genes, may turn out to have the largest number of functional genes. In this connection, it should be mentioned that the gene OR83b (OR49 in our notation) is known to be coexpressed with another OR gene in most olfactory receptor neurons (ORNs) (10, 11) and is highly conserved even among different orders of insects (12).

Yet, this gene in simulans was judged as a pseudogene because it contained a stop codon. However, if we consider the functional importance of this gene, the stop codon is likely to have occurred because of sequencing errors. We have therefore decided to regard it as a functional gene in this paper. Furthermore, the previous study showed that melanogaster has 62 functional OR genes (4), but we regarded one of them (OR83e or OR55) as a pseudogene because of large deletions.

Chromosomal Locations of OR Genes and Their Phylogenetic Relationships. Fig. 1 shows the chromosomal locations of OR genes in melanogaster, yakuba, and pseudoobscura, whose genome sequences are better assembled than others. OR genes are numbered from the left-hand side of the genome to the right-hand side in each species. In all these species, OR genes are widely distributed in the genome except for the dot chromosome (chromosome 4 in melanogaster and yakuba, and chromosome 5 in pseudoobscura), where no OR gene was observed. Phylogenetic analysis of these OR genes revealed their orthologous and paralogous relationships among the three species (see Fig. 2 for a segment of the phylogenetic tree). The orthologous genes between melanogaster and yakuba are arranged in the genome in

*Author contributions: M. Nozawa and M. Nei designed research; M. Nozawa analyzed data; and M. Nozawa and M. Nei wrote the paper.

The authors declare no conflict of interest.

Abbreviations: GO, gene order; MR, modified reconciled tree; OR, olfactory receptor; ORN, olfactory receptor neuron.

*To whom correspondence may be addressed. E-mail: mn12@psu.edu or nxm2@psu.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0702133104/DC1.

© 2007 by The National Academy of Sciences of the USA
essentially the same order. The genomic arrangement of OR
genes in pseudoobscura was quite different from that in the other
two species apparently because of the gene rearrangements that
occurred in the past. However, Fig. 2 indicates that pseudoobscura
almost always has a gene orthologous to the OR genes from the
other two species.

Comparison of the melanogaster and yakuba genomic se-
quencies shows that there are two lost genes in melanogaster
(corresponding to the open triangles in the yakuba sequence)
and one lost gene in yakuba. One melanogaster pseudogene OR55
is apparently the ortholog of yakuba gene OR55 and is likely to
be one of the two lost genes because its adjacent OR genes
showed 1:1 orthologous relationships with the same gene order.

In addition, the melanogaster genome contains one set of duplicate
genes (solid lines) that apparently occurred after divergence
of the two species. Similarly, the yakuba genome has two sets of
duplicate genes. In other words, it appears that melanogaster
acquired one gene and lost two genes whereas yakuba acquired
two genes and lost one gene after their divergence. Comparison
of the yakuba and pseudoobscura sequences suggests that yakuba
acquired four genes and lost two genes (open rectangles)
whereas pseudoobscura acquired seven genes and lost four genes
after their divergence (Fig. 1). However, these estimates may not
be reliable, because pseudoobscura and yakuba (or melanogaster)
diverged \( \sim 55 \) million years ago (MYA) (8). Therefore, we used
different methods to estimate the numbers of gene gains and
losses as shown in the next section.

The genomic maps presented in Fig. 1 are also useful for
identifying chromosomal inversion and translocation events. Genes OR61 and OR62 in melanogaster are inverted relative to
their orthologs of yakuba, and they are located on the DNA
strand opposite to that of yakuba. In addition, some genes (e.g.,
OR24 to OR27) moved from one chromosome (2R) arm to
another (2L) or vice versa. Multiple events of inversion and
translocation apparently occurred between yakuba and pseudoobscura, so that it is difficult to infer every evolutionary
event.

We noticed that 7 of the 12 duplication events observed in the
three species occurred by tandem duplication (<10 kb apart on
the same strand). This can also be seen in Fig. 2, which shows that
the genes contiguous located on the chromosomes are phylo-
genetically very close. Fig. 3 shows the relationships between the
genomic distance between two consecutive OR genes and their
phylogenetic distance (Poisson-correction distances of amino
acid sequences) within chromosomes. It is clear that there is a
high correlation between the two quantities, suggesting that
duplicate genes eventually spread through the entire chromo-
some by repeated chromosomal rearrangements.

Fig. 4 shows the NJ tree of functional OR genes from
melanogaster, pseudoobscura (subgenus Sophophora), and virilis
(subgenus Drosophila). The melanogaster gene OR83b (OR49)
and its orthologs from different species were used as outgroups,
because the gene is known to have diverged from other OR genes
a long time ago (4) and have a function different from other OR
genes (10, 11). OR genes of each species do not form a
species-specific clade but are scattered throughout the tree. We
classified OR genes into 15 phylogenetic clades (A–O), each of
which was defined as the largest cluster of similar genes sup-
ported by a bootstrap value of \( \geq 80\% \). These clades remained
unchanged even in the tree constructed for all functional OR
genes from the 12 Drosophila species [supporting information (SI)
Fig. 7]. Because all these clades contained OR genes from the
Sophophora and Drosophila species, they must have existed in the
most recent common ancestor (MRCA) of these subgenera.

The numbers of genes for the 15 phylogenetic clades in each of
the 12 Drosophila species are presented in Table 2. The number of OR genes varies considerably among these clades,
clade L having the largest number of genes. In all species, clade
O has only one gene, which is orthologous to OR33 (OR49) in
melanogaster. The genes belonging to this clade are distantly
related to other OR genes (Fig. 4). There are some other clades,
which contain only one or two genes, but it is uncertain whether
they have any distinct function or not. In this study, OR gene
groups are defined by phylogenetic relationships only, and
therefore the classification may not be related to gene function.
Yet, it is interesting to note that the number of genes in each
clade is more or less the same for all species.

### Numbers of OR Genes in Ancestral Species

We estimated the
numbers of OR genes in ancestral species and gains and losses of
genes during Drosophila evolution, using two different meth-
ods. One is the modified reconciled-tree (MR) method (17). This
method is based on the comparison of a bootstrap condensed
gene tree (18) with the species tree, and the numbers are
estimated under the parsimony principle. In this study, we used
a 50% bootstrap condensed tree of OR genes. The other is the
gene order (GO) method developed in this study. In this method,
the adjacent genes of each OR gene are considered, and when
at least one of the two adjacent genes for a pair of OR genes gives
the same best hit gene in the Blast search (19), the OR genes are
regarded as orthologs or paralogs. By using this information, the
number of gene gains and losses in the evolutionary process can
be estimated (see Materials and Methods). For this analysis, we
used representative five species including both subgenera, be-
cause the application of the GO method for a larger number of
species was complicated.

Fig. 5 shows estimates of the numbers of OR genes in ancestral
species and gene gains and losses when the five Drosophila
species were used. The estimates obtained by the MR and GO
methods are generally similar to each other. Both methods showed
that substantial numbers of gene gains and losses have
occurred for each branch of the tree. Nevertheless, both extant
and ancestral species contain similar numbers of OR genes
despite these gains and losses. Similar results were obtained
when we used different bootstrap threshold values (30–95\%) in
the MR method. These observations suggest that the OR
repertoire of Drosophila species has been quite stable in the
evolutionary process.

Comparing the genomic maps of OR genes for melanogaster
and yakuba, we previously concluded that melanogaster probably
 gained one gene and lost two genes after their divergence,
whereas yakuba gained two genes and lost one gene. This
conclusion is the same as that obtained by the MR method and
similar to that obtained by the GO method. However, compari-
on of the genomic maps of OR genes for yakuba and pseudoobscura
gave much smaller number of gene gains and losses compared with those obtained by the MR and GO methods.
These results suggest that when distantly related species are compared, the MR and GO methods give more accurate estimates than the comparison of genomic sequences.

Discussion

We have seen that the repertoire of OR genes in Drosophila species has been nearly the same for the entire period of Drosophila evolution. Yet, we identified frequent gains and losses of OR genes in the evolutionary process, indicating that the birth-and-death process (20, 21) has operated during the evolution of Drosophila OR genes. Unequal crossing over appears to be the major mechanism for increasing gene number,

Fig. 1. Chromosomal locations of OR genes and their orthologous and paralogous relationships in melanogaster (mel), yakuba (yak), and pseudoobscura (pse). Long rods show functional OR genes, whereas short rods indicate either pseudogenes or partial genes. Rods above and below the lines indicate that the OR genes are located on opposite strands. Broken and solid lines show orthologous and paralogous relationships of OR genes, respectively. Open triangles and rectangles indicate the genes whose orthologous genes were lost after the divergence of melanogaster and yakuba and after the divergence of yakuba and pseudoobscura, respectively. Stars indicate the position of OR83b (OR49) orthologs. Filled circles represent the centromeres. L and R indicate the left and right arms of chromosomes, respectively. The subscript r indicates that the chromosomal location of the sequences remains unidentified. In melanogaster and yakuba, the genes are numbered from chromosomes X to 3R, and the numbers in melanogaster are shown above the figure. In pseudoobscura, the genes are arranged from chromosomes XL to 2, because melanogaster chromosomes X, 2L, 2R, 3L, and 3R correspond to pseudoobscura chromosomes XL, 4, 3, XR, and 2, respectively (13). Species relationship is shown in the left-hand side. Three OR pseudogenes whose chromosomal locations are unidentified in pseudoobscura are not shown.

Fig. 2. Partial segment of the neighbor-joining (NJ) tree (14) of functional OR genes from melanogaster (m), yakuba (y), and pseudoobscura (p) (236 aa used). In each species, the genes are named as mentioned in Fig. 1. Filled circles indicate that the interior branches are supported by ≥80% bootstrap value. The NJ tree was constructed by using the complete deletion option of MEGA3.1 (15) after sequence alignment with CLUSTAL W (16).
because closely related genes in the phylogenetic tree were observed in tandem arrays in the genome. Although gene retroposition can increase the number of genes in a genome, this factor appears to have played negligible roles in the present case, because only two of 711 functional OR genes were intronless. Intronless genes could occur by retroposition, but they can also be generated by deletion of introns. Chromosomal rearrangements have contributed to the establishment of genome-wide distribution of OR genes, because the gene order of the orthologous OR genes is substantially different between distantly related species. These aspects of evolutionary changes of OR genes are essentially the same for both Drosophila and mammalian species (22).

However, there are two major differences in the olfactory system between mammals and Drosophila. One is the number of glomeruli in the olfactory bulb of mammals and in the antennal lobe of Drosophila relative to the number of OR genes. In Drosophila, each ORN expresses only one or two OR genes (in addition to the co-expression of gene OR83b in most neurons), and the olfactory information from ORNs, which express the same OR genes, is transmitted to a single glomerulus (23). The number of glomeruli is therefore similar to the number of different ORs. For example, melanogaster has ≈60 ORs (4) and ≈50 glomeruli (24), and the honey bee has ≈160 ORs (7) and ≈160 glomeruli (25). In mammals, each ORN also expresses only one OR gene, but the number of glomeruli is considerably greater than the number of ORs. For example, the mouse has ≈1,000 ORs (22) and 1,800 glomeruli (26), and the human has ≈400 ORs (27) and ≈8,000 glomeruli (28). These observations suggest that the olfactory information received by ORs can be transmitted to glomeruli in a more flexible way in mammals than in Drosophila.

**Table 2. Numbers of functional OR genes for 15 clades in each Drosophila species**

<table>
<thead>
<tr>
<th>Clade</th>
<th>Species*</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O†</th>
</tr>
</thead>
<tbody>
<tr>
<td>mel</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>18</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>sim</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>15</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>sec</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>15</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>yak</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>17</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ere</td>
<td>8</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>17</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ana</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>24</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>pse</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>19</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>per</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>wil</td>
<td>12</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>19</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>vir</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>17</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>moj</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>21</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>gri</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>20</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The phylogenetic clades in this table are defined in Fig. 4.
†Orthologous to the melanogaster gene OR83b (OR49).
‡Contains one internal stop codon.
The other difference is the expression pattern of OR genes. In Drosophila, a specific OR gene tends to be expressed deterministically in a specific ORN, which produces precise expression pattern of a given OR gene (29). Therefore, if an OR gene is duplicated or lost from the genome, the gene expression pattern may be disturbed. In mammals, however, one of the clustered OR genes in the genome is stochastically chosen to be expressed in each ORN (30). Therefore, the expression pattern of OR genes appears to be considerably different among different individuals, and consequently the number of OR genes may change relatively easily in the evolutionary process. Actually, at least 26 loci of OR genes are known to be polymorphic between functional and nonfunctional alleles among human individuals (31).

Of course, the most important factor for determining the number of OR genes in mammals would be environmental conditions. Mammalian species inhabit a wide range of environments, including various temperature zones, diverse habitats (terrestrial, aquatic, and aerial living), etc. It is therefore possible that a varying degree of olfaction required for living in different environments has caused the evolutionary change of OR genes and this change has occurred because of the flexible system of olfaction in mammals. Drosophila species also inhabit highly variable environments (32), but the variation in the number of OR genes may not have changed greatly because of the rigid expression and the signal pathway mentioned above.

It should be noted that mammalian species generally have a large number of OR pseudogenes. It is also known that removal of a substantial proportion of the olfactory bulb in laboratory rats does not seriously affect their survival (33). Therefore, a considerable portion of the change in the number of OR genes could be more or less neutral.

Materials and Methods

Identification of OR Genes in the Draft Genome Sequences. The genome sequences of melanogaster (release 4.3) and pseudoobscura (release 2.0) were downloaded from the FlyBase (34). Other genome sequences were obtained from the AAA database (http://rana.lbl.gov/drosophila/): simulans (dsim.washu.2jun05.mosaic.tar.gz), sechellia (dsec.broad.28oct05.tar.gz), yakuba (dyak.washu.13dec05.tar.gz), erecta (dere.freeze1.tar.gz), ananas-sue (dana.freeze1.tar.gz), persimilis (dper.broad.28oct05.tar.gz), willistonia (dwil.caf1.tar.gz), virilis (dviar.freeze1.tar.gz), mojavensis (dmoj.freeze1.tar.gz), and grimshawi (dgr1.freeze1.tar.gz).

To identify the OR genes, we performed a two-round TBlastN (19) search with E-value $\leq 10^{-10}$ against each genome sequence. In the first round, amino acid sequences of 61 OR genes annotated in the melanogaster genome were used as queries. After collecting hit sequences, we manually annotated each hit sequence, because Drosophila OR genes have introns (4) and it was not easy to determine the ORF by using computer programs. In the second round, the procedures were repeated by using the functional OR genes identified in the first round to find additional OR genes. To extract only unique genes, we compared the nucleotide sequences of all functional OR genes with one another and eliminated any overlapping genes. The flowchart for the detailed procedure is shown in SI Fig. 8. Note that the alternative transcripts from the same locus were considered as different genes. There were two such loci in the melanogaster genome. Coding sequences of functional OR genes and the genomic locations of all OR genes are available from SI Data Set 1 and SI Table 3, respectively.

Estimation of the Numbers of OR Genes in Ancestral Species. The numbers of genes in ancestral species and gains and losses of OR genes during Drosophila evolution. (a) Estimates by the MR method using a 50% condensed tree. (b) Estimates by the GO method. Numbers in circles indicate the numbers of OR genes in the extant species, whereas those in squares show the numbers in ancestral species. Numbers above and below each branch indicate the numbers of gains (+) and losses (−) of genes, respectively. The time scale shown below the tree is from ref. 8.
genes in evolution were estimated by the MR (17) and the GO methods. The latter method is expected to give more reliable results when the bootstrap support of gene trees is low, although the adjacent genes may not always be identifiable because of gene rearrangements in the past. The detailed procedure of this method is described in Fig. 6. The flowchart for the identification of adjacent genes is shown in SI Fig. 9.

We thank Saby Das, Zhenguco Lin, Jongmin Nam, Nikos Nikolaidis, Alex Rooney, Shigeru Saito, Claire T. Saito, Shozo Yokoyama, and Jianzhi Zhang for valuable comments on earlier versions of the manuscript. Steve Schaeffer provided us with information on the chromosome assembly of pseudoobscura. Jongmin Nam and Yoshihito Niimura gave us the scripts for the MR method. This work was supported by National Institutes of Health Grant GM020293 (to M. Nei).