Designing exons for human olfactory receptor gene subfamilies using a mathematical paradigm

Sk. Sarif Hassan[1,3], Pabitra Pal Choudhury[1], Amita Pal[2], R L Brahmachary[3] and Arunava Goswami[3,*]

[1]*Applied Statistics Unit,*

[2]*Bayesian Interdisciplinary Research Unit (BIRU), and*

[3]*Biological Sciences Division, Indian Statistical Institute, 203 B T Road, Calcutta 700 108, India*

[*]*Corresponding author (Email, agoswami@isical.ac.in, srabanisopanarunava@gmail.com)*

Ligands for only two human olfactory receptors are known. One of them, OR1D2, binds to Bourgeonal, a volatile chemical constituent of the fragrance of the mythical flower, Lily of the valley or Our Lady's tears, *Convallaria majalis* (also the national flower of Finland). OR1D2, OR1D4 and OR1D5 are three full-length olfactory receptors present in an olfactory locus in the human genome. These receptors are more than 80% identical in DNA sequences and have 108 base pair mismatches among them. Apparently, these mismatch positions show no striking pattern using computer pattern recognition tools. In an attempt to find a mathematical rule in those mismatches, we find that an L-system generated sequence can be inserted into the OR1D2 subfamily-specific star model and novel full-length olfactory receptors can be generated. This remarkable mathematical principle could be utilized for making new subfamily olfactory receptor members from any olfactory receptor subfamily. The aroma and electronic nose industry might utilize this rule in future.

[Hassan Sk S, Choudhury P P, Pal A, Brahmachary R L and Goswami A 2010 Designing exons for human olfactory receptor gene subfamilies using a mathematical paradigm; *J. Biosci.* **35** 389–393] **DOI** 10.1007/s12038-010-0044-0

1. Introduction

The loci of olfactory receptors (ORs) in the human genome occur in clusters ranging from ~51 to 105 and are unevenly spread over 21 chromosomes (Malnic *et al.* 2004; Young *et al.* 2008). A conservative estimate suggests that 339 full-length OR genes and 297 OR pseudogenes are present in these clusters (Malnic *et al.* 2004). Theoretically, there are two possible ways of OR-odorant molecular binding, viz. (i) each OR binds to a large number of different odorants and (ii) each OR binds to a small number of odorants. In either case, odorant detection at the OR level follows a combinatorial rule, though the stringency of the rule would differ in the two cases. Experimentally, it has been demonstrated that each OR recognizes a large number of odorants and perhaps a large class of various concentrations of the odorants tested (Malnic *et al.* 1999). The OR gene (conceptually translated to protein sequences) family (>40% amino acid identity) can be divided into subfamilies (>60% identity) and subfamily members might have more than 90% identity (Glusman *et al.* 2001). Subfamily members are highly similar in DNA and protein sequences, but they are capable of recognizing different odorant molecules.

We hypothesized that there might be a nature-inspired mathematical rule that determines the sequences of subfamily members and could be extended to subfamily and family. If such a rule exists, it would be of great interest for basic research; furthermore, one could construct ORs useful for applied research (viz. for studies in connection with an electronic nose). Three full-length model subfamily OR members were chosen from the HORDE database (*http://genome.weizmann.ac.il/horde/*), OR1D2 (gene length: 936 bp), OR1D4 (gene length: 936 bp) and OR1D5 (gene length: 936 bp). OR1D2 (chromosomal position: 17p13.3; synonym: hOR17-4) recognizes the odorant Bourgeonal which is perceived as Lily of the valley fragrance (Malnic *et al.* 2004). Incidentally, OR1D2, OR1D4 and OR1D5 show very little or no polymorphism in the published sequence

Keywords. ClustalW; human olfactory receptor; L-system; olfaction; star model

databases by different research groups (data not shown). It is possible that these groups might have used the same samples or the same source while cloning and sequencing. OR1D2, OR1D4 and OR1D5 were aligned using ClustalW and were found to contain 108 base pair mismatches out of 936 base pairs available (data not shown).

If we consider OR1D2, OR1D4 and OR1D5 each as a string of A/T/G/C, then out of 936 positions, 828 excluding 108 mismatches were found to be chosen by nature as fixed or evolutionarily conserved positions. As OR1D2, OR1D4 and OR1D5 are highly related sequences, therefore, a canonical sequence for this subfamily, termed as 'star model' of OR sequence was made by using a computer C program, where 108 gaps were introduced in the respective positions (figure 1).

A context free L-system (Prusinkiewicz and Lindenmayer 1990) was used to generate a 243 bp long DNA sequence.

L-System:

Set of variables: A, T, C and G

Axiom: C (C is the starting symbol)

Production rule: A → CTG, C→CCA, T→TGC and G→GAC

Following the production rule above, the first and second iteration would give CCA (03 bp) and CCACCACTG (09 bp), respectively. Four iterations yield 81 base pair sequences. This is insufficient to answer for 108 mismatches. Five such iterations generate the following 243 bp sequence:

CCACCACTGCCACCACTGCCATGCGACCCACC
ACTGCCACCACTGCCATGCGACCCACCACTGTGC
GACCCAGACCTGCCACCACCACTGCCACCACTGC
CATGCGACCCACCACTGCCACCACTGCCATGCGA
CCCACCACTGTGCGACCCAGACCTGCCACCACC
ACTGCCACCACTGCCATGCGACTGCGACCCAGA
CCTGCCACCACCACTGGACCTGCCACCATGCGACC
CACCACTG (i)

Using a C computer program, nucleotides present in sequence (i) were sequentially introduced from the 5′-end of the sequence into the star model gaps shown in figure 1. Briefly,

Step 1: First, in all the gaps (with 1 bp, 2 bp, 3 bp and 4 bp) in the star model, only one nucleotide would be inserted.

Step 2: 1 bp gaps in the star model would become 0 gaps. Then the remaining gaps (1 bp, 2 bp and 3 bp) would be filled up and the process would be repeated until all the gaps are filled.

The resultant OR sequence is shown in (ii) below.

ATGGATGGAGCCAACCAGAGTGAGTCCTCACAGT
TCCTTCTCCTGGGGATGTCAGAGAGTCCTGAGCAG
CAGCAGATCCTGTTTTGGATGTTCCTGTCCATGTAC
CTGGTCACGGTGCTGGGAAATGTGCTCATCATCCT
GGCCATCAGCTCTGATTCCCCCCTGCACACCCCCG
TGTACTTCTTCCTGGCCAACCTCTCCTTCACTGACC
TCTTCTTTGTCACCAACACAATCCCCAAGATGCTG
GTGAACCTCCAGTCCCAGAACAAAGCCATCTCCTA
TGCAGGGTGTCTGACACAGCTCTACTTCCTGGTCT
CCTTGGTGACCCTGGACAACCTCATCCTGGCCGTG
ATGGCCTATGATCGCTATGTGGCCAGCTGCTGCCCC
CTCCACTACGCCACAGCCATGAGCCCTGCGCTCTG
TCTCTTCCTCCTGTCCTTGTGTTGGGCGCTGTCAGT
CCTCTATGGCCTCCTGCCCACCGTCCTCATGACCAG
CGTGACCTTCTGTGGGCCTCGAGACATCCACTACG
TCTTCTGTGACATGTACCTGGTGCTGCGGTTGGCAT
GTTCCAACAGCCACATGAATCACACAGCGCTGATT
GCCACGGGCTGCTTCATCTTCCTCACTCCCTTGGGA
TTCCTGACCAGGTCCTATGTCCCCATTGTCAGACCC
ATCCTGGGAATACCCTCCGCCTCTAAGAAATACAA
AGCCTTCTCCACCTGTGCCTCCCATTTGGGTGGAG
TCTCCCTCTTATATGGGACCCTTCCTATGGTTTACCT
GGAGCCCCTCCATACCTACTCCCTGAAGGACTCAG
TAGCCACAGTGATGTATGCTGTGGTGACACCCATGA
TGAACCCGTTCATCTACAGCCTGAGGAACAAGGAC
ATGCATGGGGCTCAGGGAAGACTCCTACGCAGACC
CTTTGAGAGGCAAACA (ii)

Sequence (ii) was blasted using DNA–DNA and translated protein–protein (Blastx) search engines in the HORDE and NCBI databases from where the initial OR1D2, OR1D4 and OR1D5 sequences were obtained. Results of the Blast searches show that with the search parameters available in the HORDE website (which could not be changed by a remote user), the (ii) sequence showed 92%, 92% and 91% identity with OR1D2, OR1D4 and OR1D5, respectively. Significantly, these insertions do not produce any stop codon in the exon sequence. It is interesting to note the following rules that might govern this biological process.

(i) If one utilizes a production rule which starts with C→ CCC, then a viable OR could be produced. It is tempting to check whether the long poly C-containing region of the telomere serves as a template for insertion as in the case of DNA replication.

(ii) It seems that each OR subfamily utilizes a specific star model. We have tested the OR10J, OR10K and OR3A loci (data not shown). Rules that govern the formation of the star model for each subfamily member are in the process of analysis.

(iii) The senses of smell and taste are primordial in nature. Our current hypothesis is based on the idea that the star model or conserved region of the ORs was produced following an as yet unidentified mathematical rule quite early in evolution. Then mathematical rules such as the L-system and its

(a)

```
OR1D4__HORDE ATGGATGGAGATAACCAGAGTGAGAACTCACAGTTCCTTCTCCTGGGGAT....50
OR1D5__HORDE ATGGATGGAGATAACCAGAGTGAGAACTCACAGTTCCTTCTCCTGGGGAT
OR1D2__HORDE ATGGATGGAGGCAACCAGAGTGAAGGTTCAGAGTTCCTTCTCCTGGGGAT
             **********  ***********     *** ********************
                         12           3456   7->gaps

OR1D4__HORDE CTCAGAGAGTCCTGAGCAGCAGCAGATCCTGTTTTGGATGTTCCTGTCCA...100
OR1D5__HORDE CTCAGAGAGTCCTGAGCAGCAGCGGATCCTGTTTTGGATGTTCCTGTCCA
OR1D2__HORDE GTCAGAGAGTCCTGAGCAGCAGCGGATCCTGTTTTGGATGTTCCTGTCCA
              ********************** *************************

OR1D4__HORDE TGTACCTGGTCACGGTGCTGGGAAATGTGCTCATCATCCTGGCCATCAGC...150
OR1D5__HORDE TGTACCTGGTCACGGTGCTGGGAAATGTGCTCATCATCCTGGCCATCAGC
OR1D2__HORDE TGTACCTGGTCACGGTGGTGGGAAATGTGCTCATCATCCTGGCCATCAGC
             ****************** ***********************************

OR1D4__HORDE TCTGATTCCCACCTGCACACCCCCATGTACTTCTTCCTGGCCAACCTCTC...200
OR1D5__HORDE TCTGATTCCCACCTGCACACCCCCATGTACTTCTTCCTGGCCAACCTCTC
OR1D2__HORDE TCTGATTCCCGCCTGCACACCCCCGTGTACTTCTTCCTGGCCAACCTCTC
             ********** ************* *************************

OR1D4__HORDE CTTCACTGACCTCTTCTTTGTCACCAACACAATCCCCAAGATGCTGGTGA...250
OR1D5__HORDE CTTCACTGACCTCTTCTTTGTCACCAACACAATCCCCAAGATGCTGGTGA
OR1D2__HORDE CTTCACTGACCTCTTCTTTGTCACCAACACAATCCCCAAGATGCTGGTGA
             ************************************************

OR1D4__HORDE ACTTCCAGTCCCAGAACAAAGCCATCTCCTATGCAGGGTGTCTGACACAG...300
OR1D5__HORDE ACTTCCAGTCCCAGAACAAAGCCATCTCCTATGCAGGGTGTCTGACACAG
OR1D2__HORDE ACCTCCAGTCCCATAACAAAGCCATCTCCTATGCAGGGTGTCTGACACAG
             ** ********** ****************************************

OR1D4__HORDE CTCTACTTCCTGGTCTCCTTGGTGACCCTGGACAACCTCATCCTGGCCGT...350
OR1D5__HORDE CTCTACTTCCTGGTCTCCTTGGTGACCCTGGACAACCTCATCCTGGCCGT
OR1D2__HORDE CTCTACTTCCTGGTCTCCTTGGTGGCCCTGGACAACCTCATCCTGGCTGT
             ************************ ********************** **

OR1D4__HORDE GATGGCGTATGATCGCTATGTGGCCATCTGCTGCCCCCTCCACTATGTCA...400
OR1D5__HORDE GATGGCGTATGATCGCTATGTGGCCACCTGCTGCCCCCTCCACTATGTCA
OR1D2__HORDE GATGGCATATGACCGCTATGTGGCCATCTGCTGCCCCCTCCACTACACCA
             ****** ***** ************** ******************   **

OR1D4__HORDE CAGCCATGAGCCCTGGGCTCTGTGTCTTGCTCCTCTCCTTGTGTTGGGGG...450
OR1D5__HORDE CAGCCATGAGCCCTGGGCTCTGTGTCTTGCTCCTCTCCTTGTGTTGGGGG
OR1D2__HORDE CAGCCATGAGCCCTAAGCTCTGTATCTTACTCCTTTCCTTGTGTTGGGTC
             **************  ******* **** ***** **************

OR1D4__HORDE CTGTCTGTTCTCTATGGCCTCCTCCTCACCTTCCTCCTGACCAGGGTGAC...500
OR1D5__HORDE CTGTCTGTTCTCTATGGCCTCCTCCTCACCTTCCTCCTGACCAGGGTGAC
OR1D2__HORDE CTATCCGTCCTCTATGGCCTCATACACACCCTCCTCATGACCAGAGTGAC
             ** ** ** ************ * * **** ***** ******* *****

OR1D4__HORDE CTTCTGTGGGCCTTGAGAGATCCACTACCTCTTCTGTGACATGTACATCC...550
OR1D5__HORDE CTTCTGTGGGCCTCGAGAGATCCACTACCTCTTCTGTGACATGTACATCC
OR1D2__HORDE CTTCTGTGGGTCACGAAAAATCCACTACATCTTCTGTGAGATGTATGTAT
             ********** *  ** * ********* ********** *****  *

OR1D4__HORDE TGCTGTGGCTGGCATGTTCCAACACCCACATCATTCACACAGTGTTGATT...600
OR1D5__HORDE TGCTGTGGCTGGCATGTTCCAACACCCACATCATTCACACAGCGTTGATT
OR1D2__HORDE TGCTGAGGATGGCATGTTCCAACATTCAGATTAATCACACAGTGCTGATT
             ***** ** ***************  ** ** * ******** * *****

OR1D4__HORDE GCCACTGGCTGCTTCATCTTCCTCACCCCCTTAGGGTTCATGACCACATC...650
OR1D5__HORDE GCCACTGGCTGCTTCATCTTCCTCACCCCCTTAGGGTTCATGACCACATC
OR1D2__HORDE GCCACAGGCTGCTTCATCTTCCTCATTCCCTTTGGATTCGTGATCATTTC
             ***** ********************  ***** ** *** *** **  **

OR1D4__HORDE CTATGTACGTATTGTCAGAACCATCCTTCAAATGCCCTCAGCCTCTAAGA...700
OR1D5__HORDE CTATGTACGTATTGTCAGAACCATCCTTCAAATGCCCTCGGCCTCTAAGA
OR1D2__HORDE CTATGTGCTGATTATCAGAGCCATCCTCAGAATACCCTCAGTCTCTAAGA
             ****** *  *** ***** ********   *** ***** * ********

OR1D4__HORDE AATACAAAACCTTCTCTACCTGTGCCTCCCATTTGGGTGTGGTCTCCCTC...750
OR1D5__HORDE AATACAAAACCTTCTCTACCTGTGCCTCCCATTTGGGTGTGGTCTCCCTC
OR1D2__HORDE AATACAAAGCCTTCTCCACCTGTGCCTCCCATTTGGGTGCAGTCTCCCTC
             ******** ******* ************************  *********

OR1D4__HORDE TTTTATGGGATGCTTGCTATGGTGTACCTGCAGCCCCTCCATACCTACTC...800
OR1D5__HORDE TTTTATGGGACGCTTGCTATGGTGTACCTGCAGCCCCTCCATACCTACTC

OR1D2__HORDE TTCTATGGGACACTTTGTATGGTATACCTAAAGCCCCTCCATACCTACTC
             ** *******  ***  ****** *****  ********************

OR1D4__HORDE CATGAAGGACTCAGTAGCCACAGTGATGTATGCTGTGCTGACACCTATGA...850
OR1D5__HORDE CATGAAGGACTCAGTAGCCACAGTGATGTATGCTGTGCTGACACCTATGA
OR1D2__HORDE TGTGAAGGACTCAGTAGCCACAGTGATGTATGCTGTGGTGACACCCATGA
             ************************************* ******* ****

OR1D4__HORDE TGAACCCTTTCATCTACAGGCTGAGGAACAAAGACATGCATGGGGCTCCG...900
OR1D5__HORDE TGAACCCTTTCATCTACAGGCTGAGGAACAAAGACATGCATGGGGCTCCG
OR1D2__HORDE TGAATCCCTTCATCTACAGCCTGAGGAACAAGGACATGCATGGGGCTCTG
             **** ** ************ ************ ***************** *


OR1D4__HORDE GGAAGAGTCCTATGGAGACCCTTTCAGAGGCCTAAA...936
OR1D5__HORDE GGAAGAGTCCTATGGAGACCCTTTCAGAGGCCTAAA
OR1D2__HORDE GGAAGACTCCTAGATAAACACTTTAAGAGGCTGACA
             ****** *****   * ** **** ******  * *
                                        gaps->108
```

(b)

```
ATGGATGGAG--AACCAGAGTGA----TCA-AGTTCCTTCTCCTGGGGAT....50
-TCAGAGAGTCCTGAGCAGCAGC-GATCCTGTTTTGGATGTTCCTGTCCA...100
TGTACCTGGTCACGGTG-TGGGAAATGTGCTCATCATCCTGGCCATCAGC...150
TCTGATTCCC-CCTGCACACCCCC-TGTACTTCTTCCTGGCCAACCTCTC...200
CTTCACTGACCTCTTCTTTGTCACCAACACAATCCCCAAGATGCTGGTGA...250
AC-TCCAGTCCCA-AACAAAGCCATCTCCTATGCAGGGTGTCTGACACAG...300
CTCTACTTCCTGGTCTCCTTGGTG-CCCTGGACAACCTCATCCTGGC-GT...350
GATGGC-TATGA-CGCTATGTGGCCA-CTGCTGCCCCCTCCACTA---CA...400
CAGCCATGAGCCCT--GCTCTGT-TCTT-CTCCT-TCCTTGTGTTGGG--...450
CT-TC-GT-CTCTATGGCCTC-T-C-CACC-TCCTC-TGACCAG-GTGAC...500
CTTCTGTGGG-C--GA-A-ATCCACTAC-TCTTCTGTGA-ATGTA--T--...550
TGCTG-GG-TGGCATGTTCCAACA--CA-AT-A-TCACACAG-G-TGATT...600
GCCAC-GGCTGCTTCATCTTCCTCA--CCCTT-GG-TTC-TGA-CA--TC...650
CTATGT-CG-ATT-TCAGA-CCATCCT---AAT-CCCTC-G-CTCTAAGA...700
AATACAAA-CCTTCTC-ACCTGTGCCTCCCATTTGGGTG--GTCTCCCTC...750
TT-TATGGGA--CTT--TATGGT-TACCT--AGCCCCTCCATACCTACTC...800
--TGAAGGACTCAGTAGCCACAGTGATGTATGCTGTG-TGACACC-ATGA...850
TGAA-CC-TTCATCTACAG-CTGAGGAACAA-GACATGCATGGGGCTC-G...900
GGAAGA-TCCTA---A-AC-CTTT-AGAGGC--A-A...936
```

Figure 1. **(a).** ClustalW of three full-length OR sequences (OR1D4, OR1D5 and OR1D2) of the the OR1D subfamily locus as found from the HORDE database. Asterisks and gaps numbered with numerals below the sequence demonstrate the conserved and variable base pairs, respectively. **(b).** Star model of the OR1D subfamily of OR gene sequences generated based on data from figure 1a.

variants were used to make the variable regions which contribute to the odorant ligand-binding domains of the ORs. This process of insertion might have happened at the DNA polymerization level.

We have already mentioned earlier in the text that there are 2–5 highly related yet diverse OR subfamily sequences clustered in the human genome. The reason and significance of this special genomic architectural plan has to be searched for in an evolutionary framework at the theoretical level. The results obtained following the aforesaid production rule as spelt out tempt us to test the hypothesis – whether nature follows this procedure or not. A comparative study of the usage of L-systems in the olfactory subgenomes of lower vertebrates such as mouse with that of human might offer clues in this direction.

In summary, in this paper, we report a relatively simple model of a context-free L-system for making a variable region of the OR and this could be adopted for making artificial ORs. Many more advanced context-free L-systems could be designed once it is experimentally established that this is the kind of rule the OR utilizes for generating subfamily members, at least, if not subfamily and family members, more divergent ORs in the genome. Here, we observe that the computer-generated star model sequence, sequentially filled with A, T, G, C in the way described above from a sequence generated by an L-system could generate a sequence that is highly similar to those of OR1D2, OR1D4 and OR1D5. Therefore, most likely, this work is purely mathematical in nature at this stage and a large body of experimental evidence is necessary.

Acknowledgements

This work was supported by the Department of Biotechnology (DBT), New Delhi, grants (BT/PR9050/NNT/28/21/2007 and BT/PR8931/NNT/28/07/2007 to AG) and NAIP-ICAR-World Bank grant (Comp-4/C3004/2008-09; Project leader: AG) and ISI plan projects for 2001–2011. The authors are grateful to their visiting students Rajneesh Singh, Snigdha Das and Somnath Mukherjee for their technical help in making advanced C programs and other computer applications on Windows support used for this study.

Appendix

While writing the computer program, the L-system satisfied following rules. (A) To fill the single gap of the star model: the system will check the two previous states as well as the two past states of the gap. (a) If the second previous is ‘T’ and the first previous is ‘A’, and the first past is ‘A’ and the second past is either ‘A’ or ‘G’, then the chosen L-system must produce ‘C’ at the gap, e.g. …TA_AA(/G)… (b) If the second previous is ‘T’ and the first previous is ‘A’ , the first past is ‘G’ and the second past is ‘A’, then the chosen L-system must produce ‘C’ at the gap, e.g.,…TA_GA… (c) If the second previous is ‘T’ and the first previous is ‘A’, and the first past is either ‘T’ or ‘C’, then the chosen L-system must produce ‘C’ or ‘T’ at the gap, e.g. …TA_T(/C)… (d) If the second previous is ‘T’ and the first previous is ‘G’, and the first past is ‘A’ and the second past is either ‘A’ or ‘G’, then the chosen L-system must produce ‘C’ or ‘G’ at the gap, e.g. …TG_AA(/G)… (e) If the second previous is ‘T’ and the first previous is ‘G’, and the first past is ‘G’ and the second past is ‘A’, then the chosen L-system must produce ‘C’ or ‘G’ at the gap, e.g. …TG_GA… (f) If the second previous is ‘T’ and the first previous is ‘G’, and the first past is either ‘T’ or ‘C’, then the chosen L-system must produce ‘C’ or ‘G’ or ‘T’ at the gap, e.g. …TG_T(/C)… (g) If the first previous is ‘T’, and the first past is ‘A’ and the second past is either ‘C’ or ‘T’, then the chosen L-system must produce ‘T’ or ‘C’ at the gap, e.g. …T_AC(/T)… (h) If the first previous is ‘T’, and the first past is ‘A’ and the second past is either ‘A’ or ‘G’, then the chosen L-system must produce ‘C’ at the gap, e.g. …T_AA(/G)… (i) If the first previous is ‘T’, and the first past is ‘G’ and the second past is ‘A’, then the chosen L-system must produce ‘G’ or ‘C’ at the gap, e.g. …T_GA… (j) If the first previous is ‘T’, and the first past is ‘G’ and the second past is ‘C’ or ‘T’ or ‘G’, then the chosen L-system must produce ‘T’ or ‘C’ or ‘G’ at the gap, e.g. …T_GC(/T/G)… (h) If the first previous is ‘C’, and the first past is ‘A’ and the second past is either ‘A’ or ‘G’, then the chosen L-system must produce ‘C’ or ‘G’ or ‘A’ at the gap, e.g. …C_AA(/G)… (i) If the first previous is ‘C’, and the first past is ‘G’ and the second past is ‘A’, then the chosen L-system must produce ‘C’ or ‘G’ or ‘A’ at the gap, e.g. …C_GA… (j) Else the gap can be filled by any state such as ‘A’ or ‘C’ or ‘T’ or ‘G’. (B) To fill the double or more than double gap of the star model (one gap fill at a time): the system should check only two previous states of the gap. (a) If the second previous is ‘T’ and the first previous is ‘A’, then the chosen L-system must produce ‘C’ or ‘T’ at the gap, e.g. …TA_ _ … (b) If the second previous is ‘T’ and the first previous is ‘G’, then the chosen L-system must produce ‘C’ or ‘T’ or ‘G’ at the gap, e.g. …TG_ _… (c) Else the gap can be filled by any state such as ‘A’ or ‘C’ or ‘T’ or ‘G’. This rule would be applicable until the number of gaps becomes one. When the number of gaps becomes one, then rule (A) is applicable.

References

Glusman G, Yanai I, Rubin I and Lancet D 2001 The complete human olfactory subgenome; *Genome Res.* **11** 685–702

Malnic B, Godfrey P-A and Buck L-B 2004 The human olfactory receptor gene family; *Proc. Natl. Acad. Sci. USA* **101** 2584–2589 Erratum in: *Proc. Natl. Acad. Sci. USA* 2004 **101** 7205

Malnic B, Hirono J, Sato T and Buck L-B 1999 Combinatorial receptor codes for odors; *Cell* **96** 713–723

Prusinkiewicz P and Lindenmayer A 1990 in *The algorithmic beauty of plants* (New York: Springer-Verlag)

Young J-M, Endicott R-M, Parghi S-S, Walker M, Kidd J-M and Trask B-J 2008 Extensive copy-number variation of the human olfactory receptor gene family; *Am. J. Hum. Genet.* **83** 228–242

MS received 19 January 2010; accepted 3 May 2010

*e*Publication: 18 June 2010

Corresponding editor: SHAHID KHAN