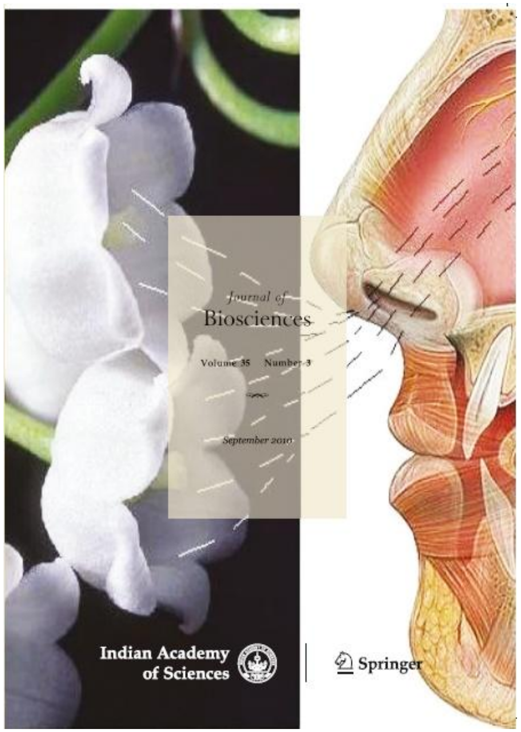
Pattern in The Organic World: From Nautilus Shell to L-System in Genome



Journal of

Biosciences

Volume 35 Number 3

September 2010

Indian Academy of Sciences

Springer

Prof. R. L. Brahmachary will deliver an introduction.

Sk. Sarif Hassan
Applied Statistics Unit

Indian Statistical Institute

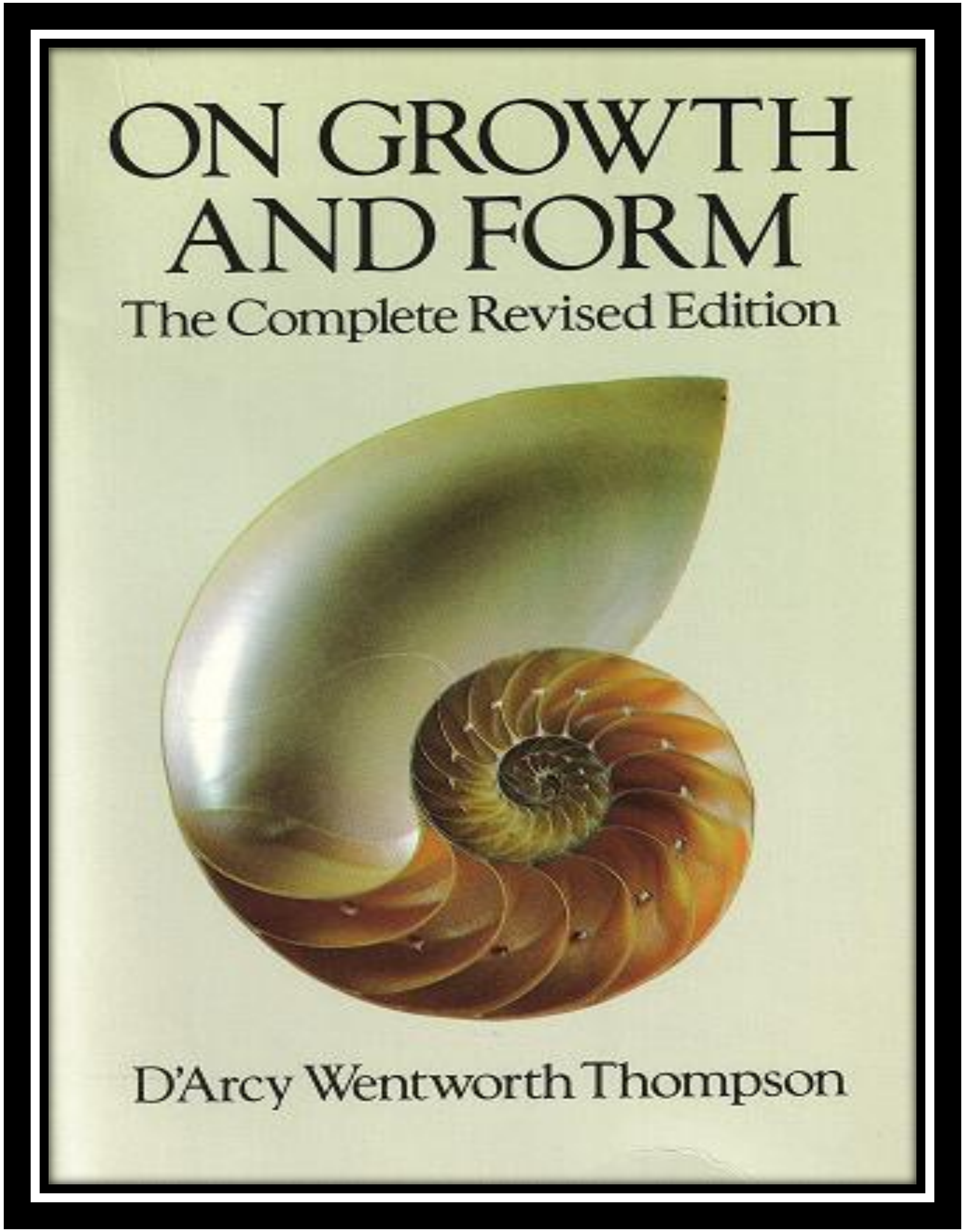
Kolkata

The beauty of symmetric (and asymmetric) pattern in the living world, as in flower or a nautilus shell has been noticed by people since the earliest times.

In 1916, D’arcy Thomson, a biologist initiated his classic work, On Growth and Form, the revised edition of which was published in 1942 and in more than a 1000 pages he treated symmetry and also ordered asymmetry in animals and plants like leaves, flower, shells, horns etc.

Simple mathematics and geometric representations form the framework of a broad perspective for viewing the well ordered beauty in the structures and growth patterns of organisms.

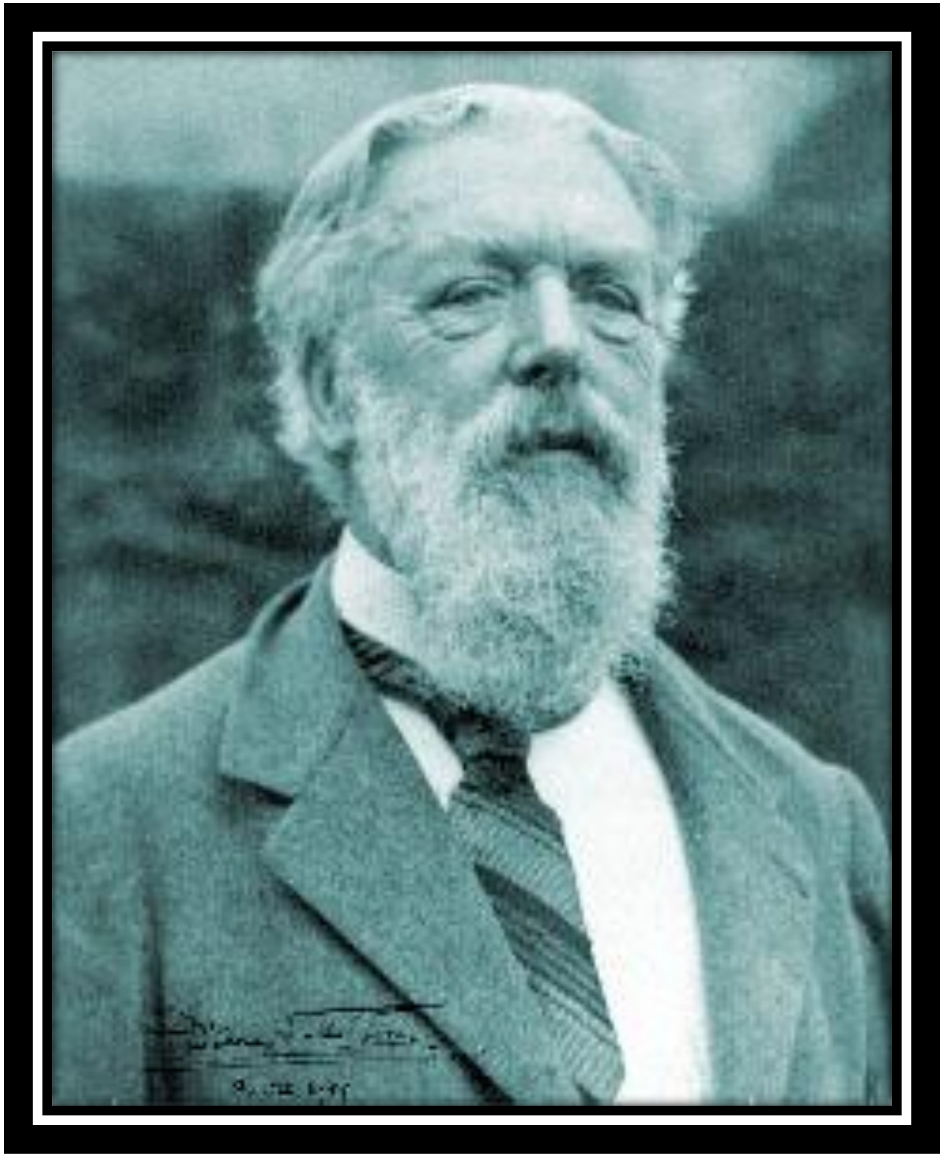
Thompson said, ---" It is not the biologist with an inkling of mathematics but the skilled & learned mathematicians who must ultimately deal with such problems as are sketched and adumbrated here".



ON GROWTH AND FORM

The Complete Revised Edition

D'Arcy Wentworth Thompson



D'Arcy Wentworth Thompson CB FRS FRSE (2 May 1860, Edinburgh – 21 June 1948, St Andrews) was a Scottish biologist, mathematician, and classics scholar.

Nautilus Shell

$$r = ae^{b\theta} \text{ where } a, b \text{ are parameters.}$$

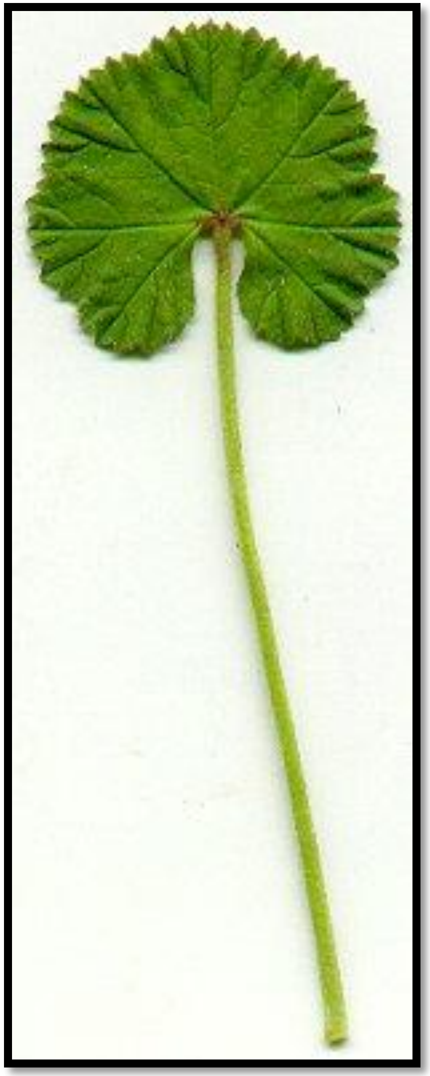
Nautilus is a marine animal (mollusc) distantly related to octopus. Unlike the latter which can be found in shallow water and is very familiar, Nautilus inhabits the deep seas. Generally we notice the shells of dead Nautilus cast on the sea beach.

With progressive age each chamber becomes larger than the previous one, always bearing a content proportion to the former.

Interesting fact is the proportion is basically the **Golden ratio**

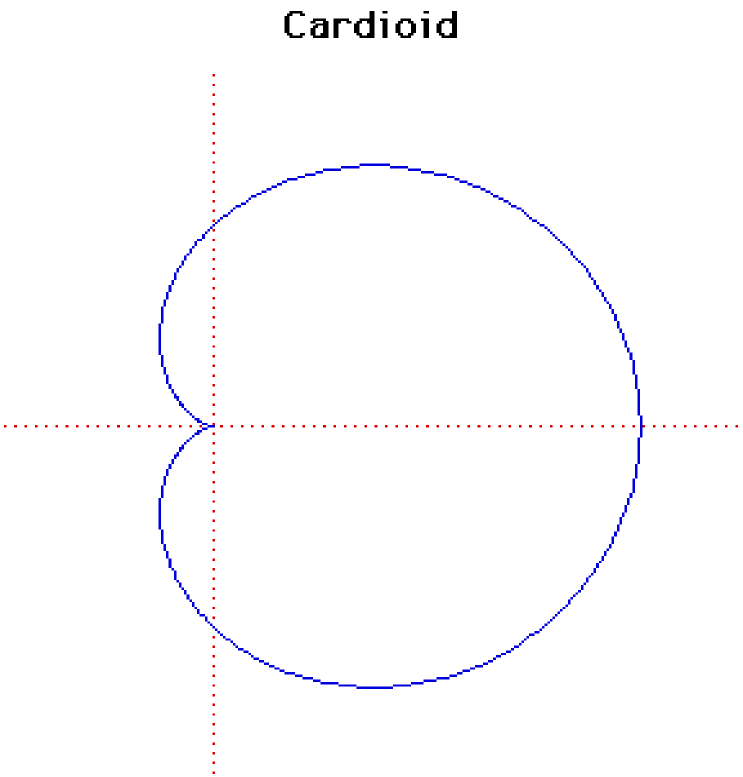
Likewise, Thompson described a curve resembling the outline of a reniform leaf. Variations of the formula r=sin(ϕ/2) can generate various shapes.

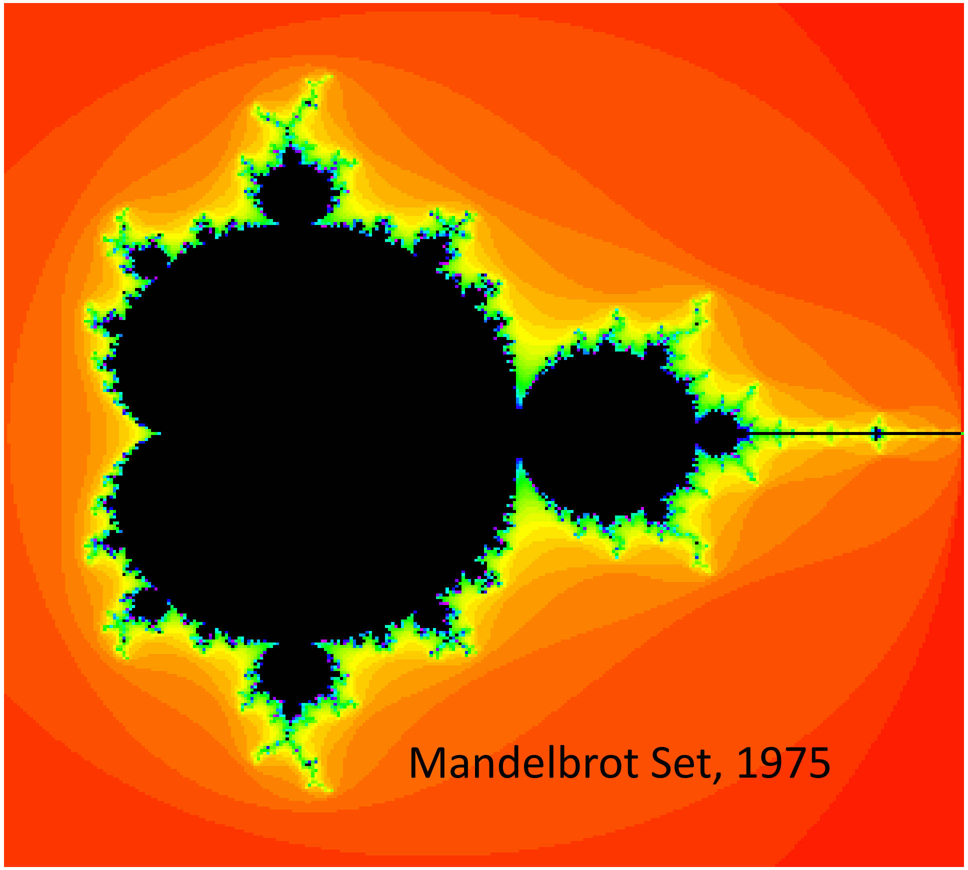
The late T. A. Davis, ISI, worked on Sunflower head.

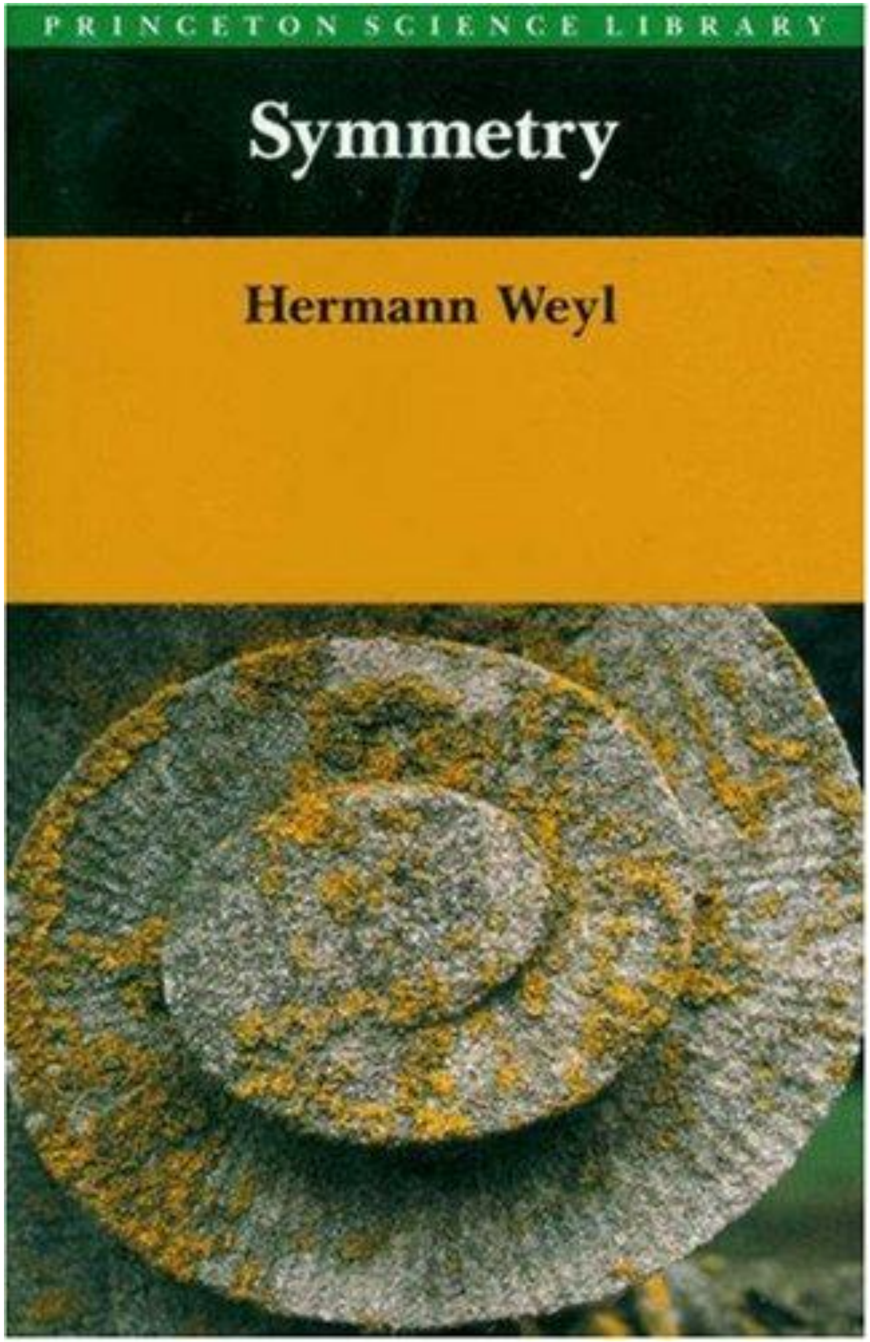


The symmetry pattern in the spirals of the sunflower head was noticed by Thompson. Around 1970, T. A. Davis of the Indian Statistical Institute worked on this aspect. Some details are sketched below:

Thompson points out that as early as in 1728 Grandi developed a class of mathematical curves and pointed out the botanical analogies.



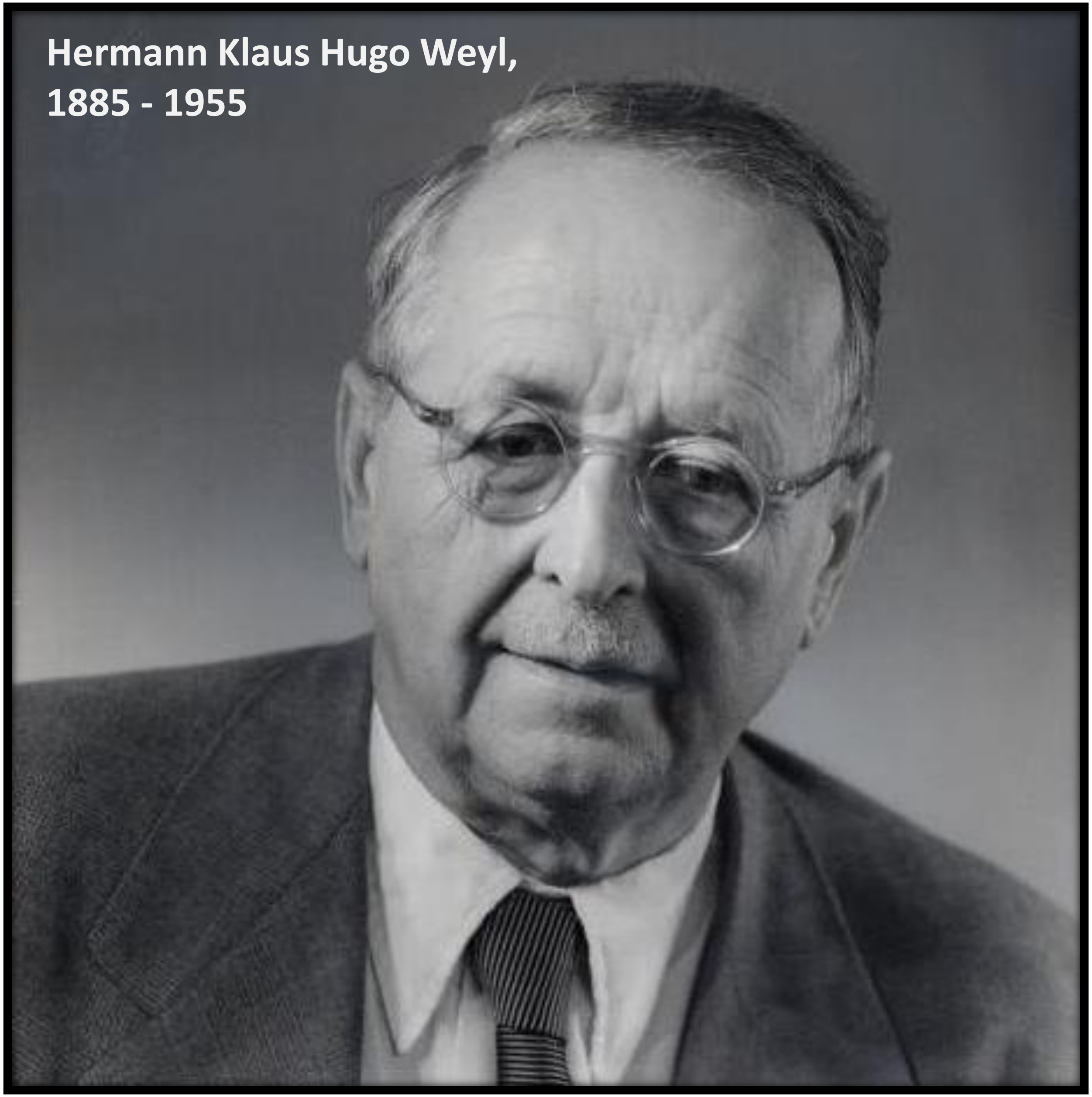




PRINCETON SCIENCE LIBRARY

Symmetry

Hermann Weyl



Hermann Klaus Hugo Weyl,
1885 - 1955

In 1951, **Hermann Klaus Hugo Weyl**, a great early exponent of Relativity theory and, Group Theory & Quantum mechanics and also wrote a book, SYMMETRY.

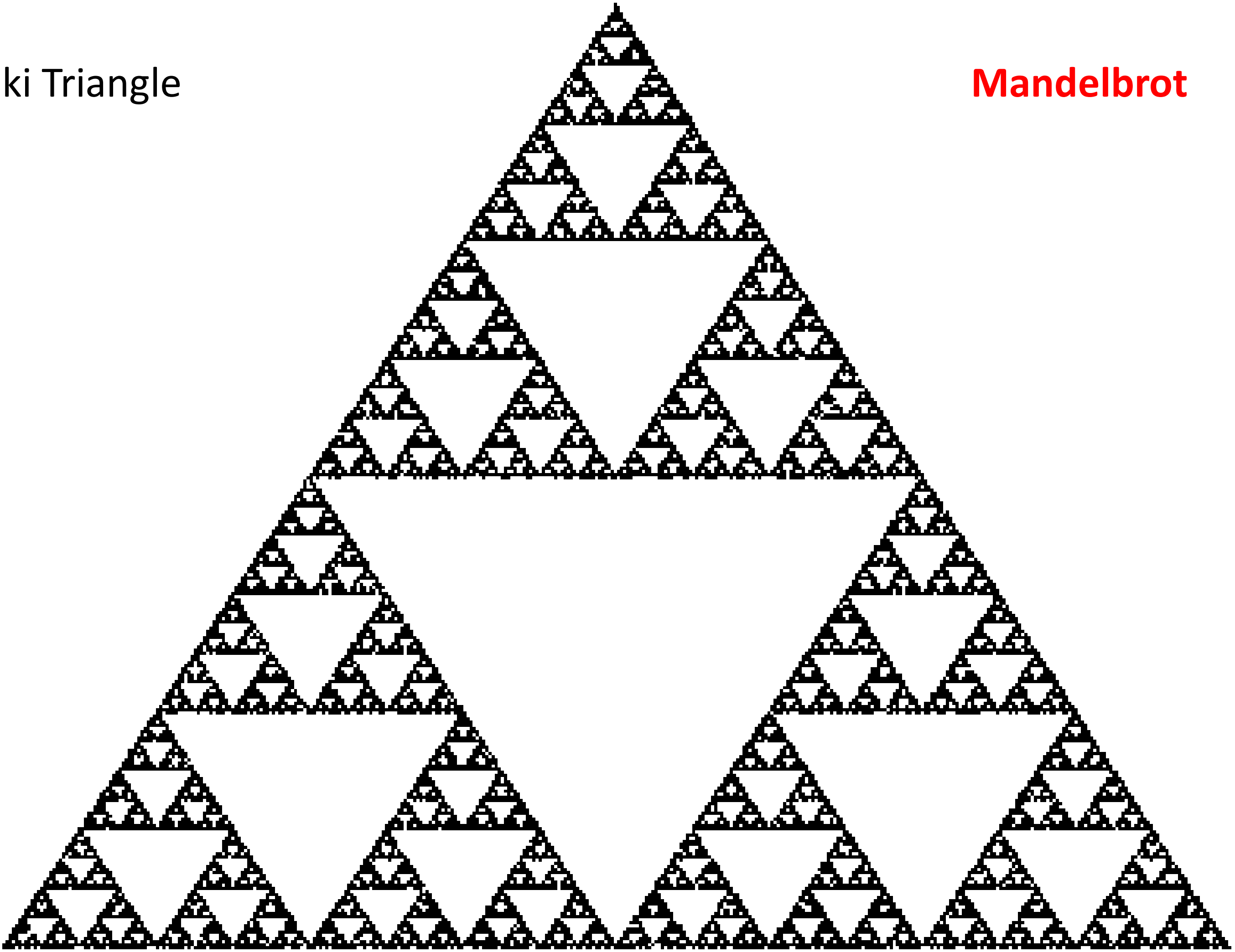
In this work, Weyl refers to D'arcy Thompson and draws from many other sources such as Haeckel's collection of drawings on diatoms (unicellular organism with beautiful silica shell), brittle stars (echinodermata) and medusae. Many of these have bilateral, radial and/or spherical symmetry. He also emphasized the surprisingly symmetrical figures of ice crystals.

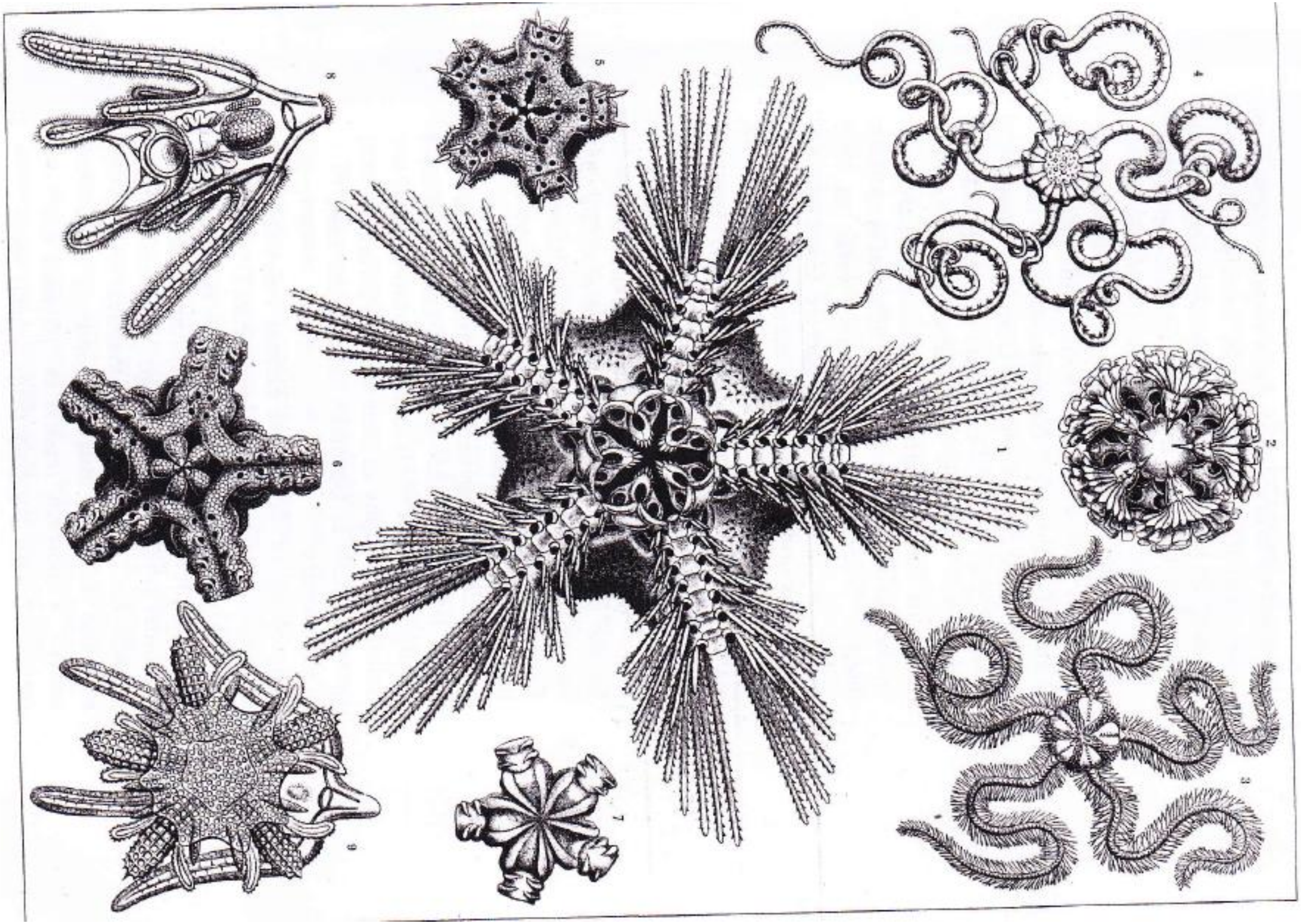
Mathematicians like B. Mandelbrot (1975) have been interested in FRACTALS, i.e. pattern of self similarity which was also pointed out by Weyl.

Self Symmetry

Sierpinski Triangle

Mandelbrot

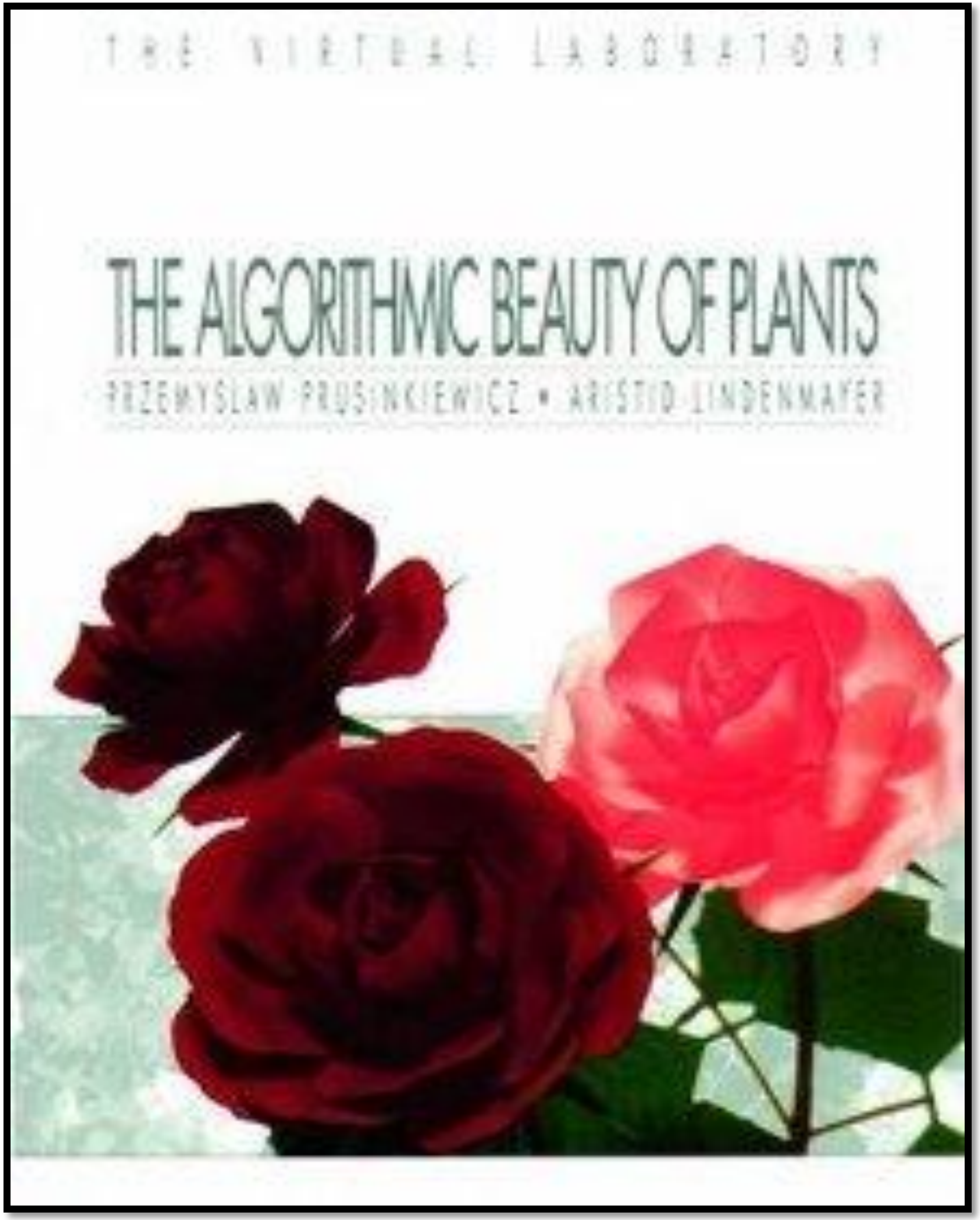



8
5
4
9
1
3
6
7
2

In metameric structural patterns in animals such as in the centipede shown here the different segments of the body are nearly of the same size. Likewise in plants buds or branchlets may appear at nearly regular distances in the branches. Weyl points out that such regular intervals translated into one dimensional time would be equivalent to a rhythm in music.

The mathematics of music is not yet developed and he says it is no wonder because the mathematics of symmetry group was developed some four thousand years after such symmetry was depicted in Egyptian Art.

More relevant for our purpose, namely to study the underlying mathematical principle of genome sequences, is the Lindenmayer System. In short, L-System.



THE VIRTUAL LABORATORY

THE ALGORITHMIC BEAUTY OF PLANTS

PRZEMYSLAW PRUSINKIEWICZ • ARISTID LINDENMAYER

Aristid Lindenmayer (November 17, 1925 – October 30, 1989) was a Hungarian biologist

Lindenmayer Systems (L-systems)

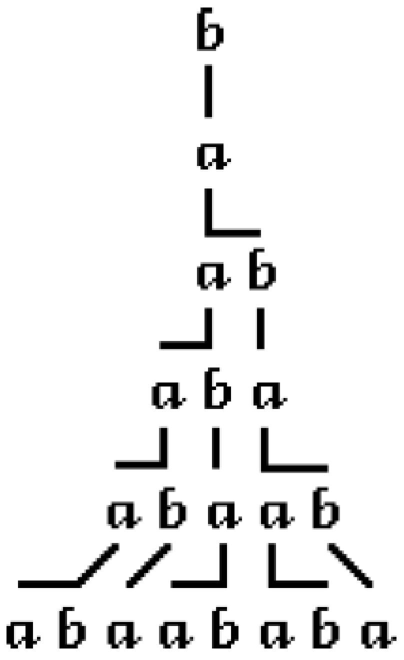
The central concept of L-systems is that of *Rewriting*. In general, *Rewriting* is a technique for defining complex objects by successively replacing parts of a simple initial object using a set of *rewriting rules or productions.*

Consider strings built of two letters *a and b*,

Production Rules:

a* → *ab

b* → *a



Example of a derivation in a DOL-system

We have started our journey with Human Olfactory Receptor OR1D2.

*** Why Human?**

****Why Olfactory Receptor?**

*****Why OR1D2?**

Why Human?

- **Human genome has been sequenced.**

*The **Human Genome Project (HGP)** was an international scientific research project with a primary goal to determine the sequence of chemical base pairs which make up DNA and to identify and map the approximately 20,000–25,000 genes of the human genome from both a physical and functional standpoint.

**Geneticists of Leiden University Medical Centre (LUMC) are the first to determine the DNA sequence of a woman. She is also the first European whose DNA sequence has been determined.

The results will contribute to insights into human genetic diversity.

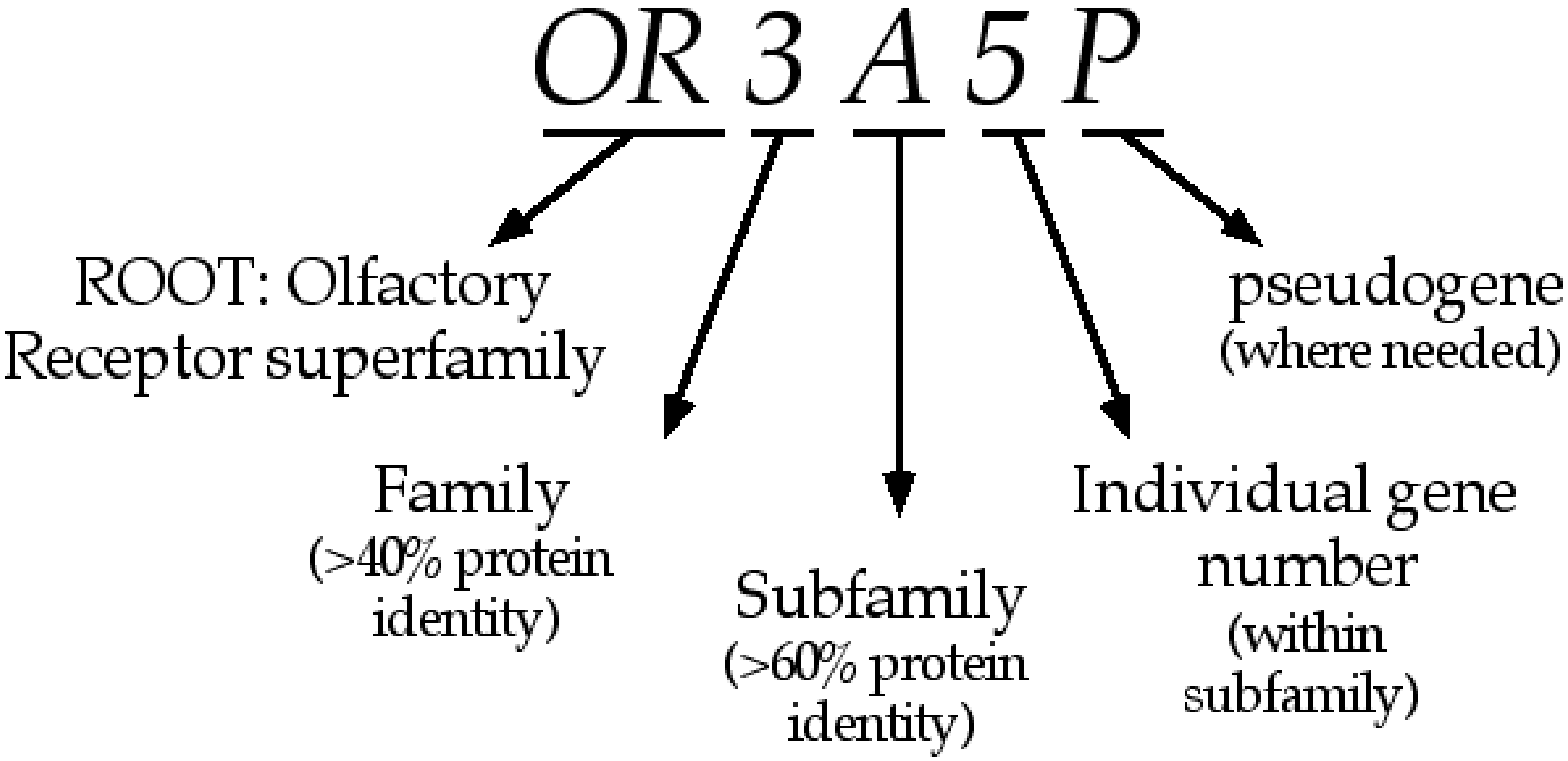
Why we start our journey with Olfaction

We took Olfactory Receptors (ORs) because of their unique features…

- ORs loci in human genome occur in clusters and are unevenly spread over 21 chromosomes.
- Human ORs are free from any stop codon. Such receptors are known to the biologist as Exons.
- OR sequence length is relatively small, almost 1000bp.

Before going to ‘Why OR1D2’ let us warm up with OR1D2

The nomenclature terms are accepted as official gene symbols by the HUGO Gene Nomenclature Committee (HGNC)



OR 3 A 5 P

ROOT: Olfactory Receptor superfamily

Family (>40% protein identity)

Subfamily (>60% protein identity)

Individual gene number (within subfamily)

pseudogene (where needed)

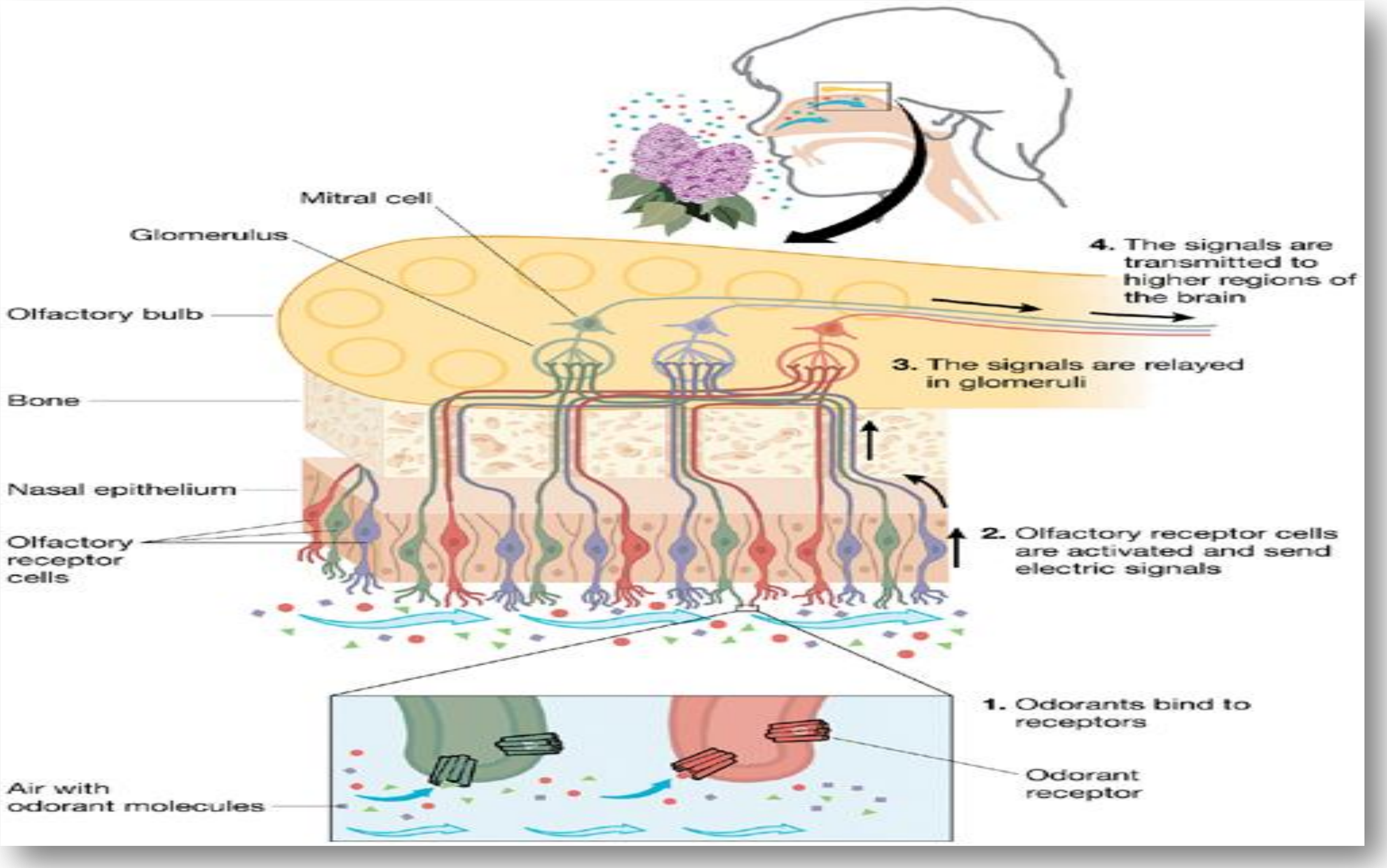
Proposed Nomenclature for Olfactory Receptor Genes

WHY OR1D2?

- **Ligands for only Twelve human olfactory receptors are known.**

- One of them, OR1D2, binds to Bourgeonal, a volatile chemical constituent of the fragrance of **Lily of the valley** **Convallaria majalis** (also the **national flower of Finland).**

A Human OR


Mitral cell
Glomerulus
Olfactory bulb
Bone
Nasal epithelium
Olfactory receptor cells
Air with odorant molecules
4. The signals are transmitted to higher regions of the brain
3. The signals are relayed in glomeruli
2. Olfactory receptor cells are activated and send electric signals
1. Odorants bind to receptors
Odorant receptor

>Human Olfactory Receptor (OR1D2): Full length gene sequence (Exon)= 936 bp (A,T,G,C)

**ATGGATGGAGGCAACCAGAGTGAAGGTTCAGAGTTCCTTCTCCTGGGGATGTCAGAGAGTC
CTGAGCAGCAGCGGATCCTGTTTTGGATGTTCCTGTCCATGTACCTGGTCACGGTGGTGGG
AAATGTGCTCATCATCCTGGCCATCAGCTCTGATTCCCGCCTGCACACCCCCGTGTACTTC
TTCCTGGCCAACCTCTCCTTCACTGACCTCTTCTTTGTCACCAACACAATCCCCAAGATGC
TGGTGAACCTCCAGTCCCATAACAAAGCCATCTCCTATGCAGGGTGTCTGACACAGCTCTA
CTTCCTGGTCTCCTTGGTGGCCCTGGACAACCTCATCCTGGCTGTGATGGCATATGACCGC
TATGTGGCCATCTGCTGCCCCCTCCACTACACCACAGCCATGAGCCCTAAGCTCTGTATCT
TACTCCTTTCCTTGTGTTGGGTCCTATCCGTCCTCTATGGCCTCATACACACCCTCCTCAT
GACCAGAGTGACCTTCTGTGGGTCACGAAAAATCCACTACATCTTCTGTGAGATGTATGTA
TTGCTGAGGATGGCATGTTCCAACATTCAGATTAATCACACAGTGCTGATTGCCACAGGCT
GCTTCATCTTCCTCATTCCCTTTGGATTCGTGATCATTTCCTATGTGCTGATTATCAGAGC
CATCCTCAGAATACCCTCAGTCTCTAAGAAATACAAAGCCTTCTCCACCTGTGCCTCCCAT
TTGGGTGCAGTCTCCCTCTTCTATGGGACACTTTGTATGGTATACCTAAAGCCCCTCCATA
CCTACTCTGTGAAGGACTCAGTAGCCACAGTGATGTATGCTGTGGTGACACCCATGATGAA
TCCCTTCATCTACAGCCTGAGGAACAAGGACATGCATGGGGCTCTGGGAAGACTCCTAGAT
AAACACTTTAAGAGGCTGACA**

***** Sequence is collected from HORDE**
http://genome.weizmann.ac.il/horde/

The Expasy Resulted Protein code for OR1D2 which is free from stop codon

- **Met** D G G N Q S E G S E F L L L G **Met** S E S P E Q Q R I L F W **Met** F L S **Met** Y L V T V V G N V L I I L A I S S D S R L H T P V Y F F L A N L S F T D L F F V T N T I P K **Met** L V N L Q S H N K A I S Y A G C L T Q L Y F L V S L V A L D N L I L A V **Met** A Y D R Y V A I C C P L H Y T T A **Met** S P K L C I L L L S L C W V L S V L Y G L I H T L L **Met** T R V T F C G S R K I H Y I F C E **Met** Y V L L R **Met** A C S N I Q I N H T V L I A T G C F I F L I P F G F V I I S Y V L I I R A I L R I P S V S K K Y K A F S T C A S H L G A V S L F Y G T L C **Met** V Y L K P L H T Y S V K D S V A T V **Met** Y A V V T P **Met** **Met** N P F I Y S L R N K D **Met** H G A L G R L L D K H F K R L T

An Invitation To The Problem

It appears that OR1D2 is a string of A, T, C, and G RANDOMLY inserted one after another.

In realty, is it the case???

Had it been so, then neither we would be able to smell the fragrance of rose properly nor distinguish it from that of Lilly.

So, There Should Be A Beautiful Organization in The Sequence.

Origin of work

Nature could make 4**936 OR sequences, instead nature has selected only 300 (Approx) as functional OR in human and 300 are considered to be non-functional (called pseudo genes).

Now, what governs this selection process?

What are the selection keys?

OR1D2, OR1D4 and OR1D5

- **OR1D2 was used as query in BLASTn search for similar ORs in HORDE and OR1D4 and OR1D5 with more than 80% identical sequences were found in the same genomic loci and the above three ORs have 108 base pair mismatches among them.**

Methodology: L-System

- In an attempt to find a mathematical rule in those mismatches, we find that L-system generated sequence can be inserted into the OR1D2 subfamily specific **Star-Model** and a close relative of the full length olfactory receptors of the same subfamily can be generated.

The 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

The L-System Generated Sequence

**ATGGATGGAGCCAACCAGAGTGAGTCCTCACAGTTCCTTCTCCTGGGGATGTCAGAGAGTCC
TGAGCAGCAGCAGATCCTGTTTTGGATGTTCCTGTCCATGTACCTGGTCACGGTGCTGGGAA
ATGTGCTCATCATCCTGGCCATCAGCTCTGATTCCCCCCTGCACACCCCCGTGTACTTCTTCC
TGGCCAACCTCTCCTTCACTGACCTCTTCTTTGTCACCAACACAATCCCCAAGATGCTGGTGA
ACCTCCAGTCCCAGAACAAAGCCATCTCCTATGCAGGGTGTCTGACACAGCTCTACTTCCTG
GTCTCCTTGGTGACCCTGGACAACCTCATCCTGGCCGTGATGGCCTATGATCGCTATGTGGCC
AGCTGCTGCCCCCTCCACTACGCCACAGCCATGAGCCCTGCGCTCTGTCTCTTCCTCCTGTCC
TTGTGTTGGGCGCTGTCAGTCCTCTATGGCCTCCTGCCCACCGTCCTCATGACCAGCGTGACC
TTCTGTGGGCCTCGAGACATCCACTACGTCTTCTGTGACATGTACCTGGTGCTGCGGTTGGCA
TGTTCCAACAGCCACATGAATCACACAGCGCTGATTGCCACGGGCTGCTTCATCTTCCTCACT
CCCTTGGGATTCCTGACCAGGTCCTATGTCCCCATTGTCAGACCCATCCTGGGAATACCCTCC
GCCTCTAAGAAATACAAAGCCTTCTCCACCTGTGCCTCCCATTTGGGTGGAGTCTCCCTCTTA
TATGGGACCCTTCCTATGGTTTACCTGGAGCCCCTCCATACCTACTCCCTGAAGGACTCAGTA
GCCACAGTGATGTATGCTGTGGTGACACCCATGATGAACCCGTTCATCTACAGCCTGAGGAA
CAAGGACATGCATGGGGCTCAGGGAAGACTCCTACGCAGACCCTTTGAGAGGCAAACA**

The result from Expasy

- **Met** D G A N Q S E S S Q F L L L G **Met** S E S P E Q Q Q I L F W **Met** F L S **Met** Y L V T V L G N V L I I L A I S S D S P L H T P V Y F F L A N L S F T D L F F V T N T I P K **Met** L V N L Q S Q N K A I S Y A G C L T Q L Y F L V S L V T L D N L I L A V **Met** A Y D R Y V A S C C P L H Y A T A **Met** S P A L C L F L L S L C W A L S V L Y G L L P T V L **Met** T S V T F C G P R D I H Y V F C D **Met** Y L V L R L A C S N S H **Met** N H T A L I A T G C F I F L T P L G F L T R S Y V P I V R P I L G I P S A S K K Y K A F S T C A S H L G G V S L L Y G T L P **Met** V Y L E P L H T Y S L K D S V A T V **Met** Y A V V T P **Met** **Met** N P F I Y S L R N K D **Met** H G A Q G R L L R R P F E R Q T

Conclusion in Research

We claim In two ways:

1. ***If The above sequence functions*** **as OR1D2, then we could say that this remarkable mathematical principle could be utilized for making new subfamily OR members from any OR subfamily. Aroma and electronic nose industry might utilize this rule in future.**
2. ***If the above sequence is not at all functioning*** **as OR1D2 then, in spite of 92% similarity, it does not yield the functionality as OR1D2. Thus actual functional part or selection key must be with the remaining 8%.**

A Big Question to The Biologists!!

- **Since already we know the functioning of OR1D2, our research outcome has to be experimentally confirmed about the dilemma as posed in the previous slide.**
- **That is, whether the L-System designed DNA sequence is at all functioning.**

Another Approach

- Now you may have the following question…
- We have designed a variable region of ORs. Why not the full receptors?
- To answer this question, we are ready with a set of L-systems to capture the whole receptor even genome.
- The procedure is shown in the next few slides…

A New principle

But, Does the Procedure have some novelty?

The beauty of the proposed methodology might enable us to answer some unanswered questions in "Evolutionary Science".

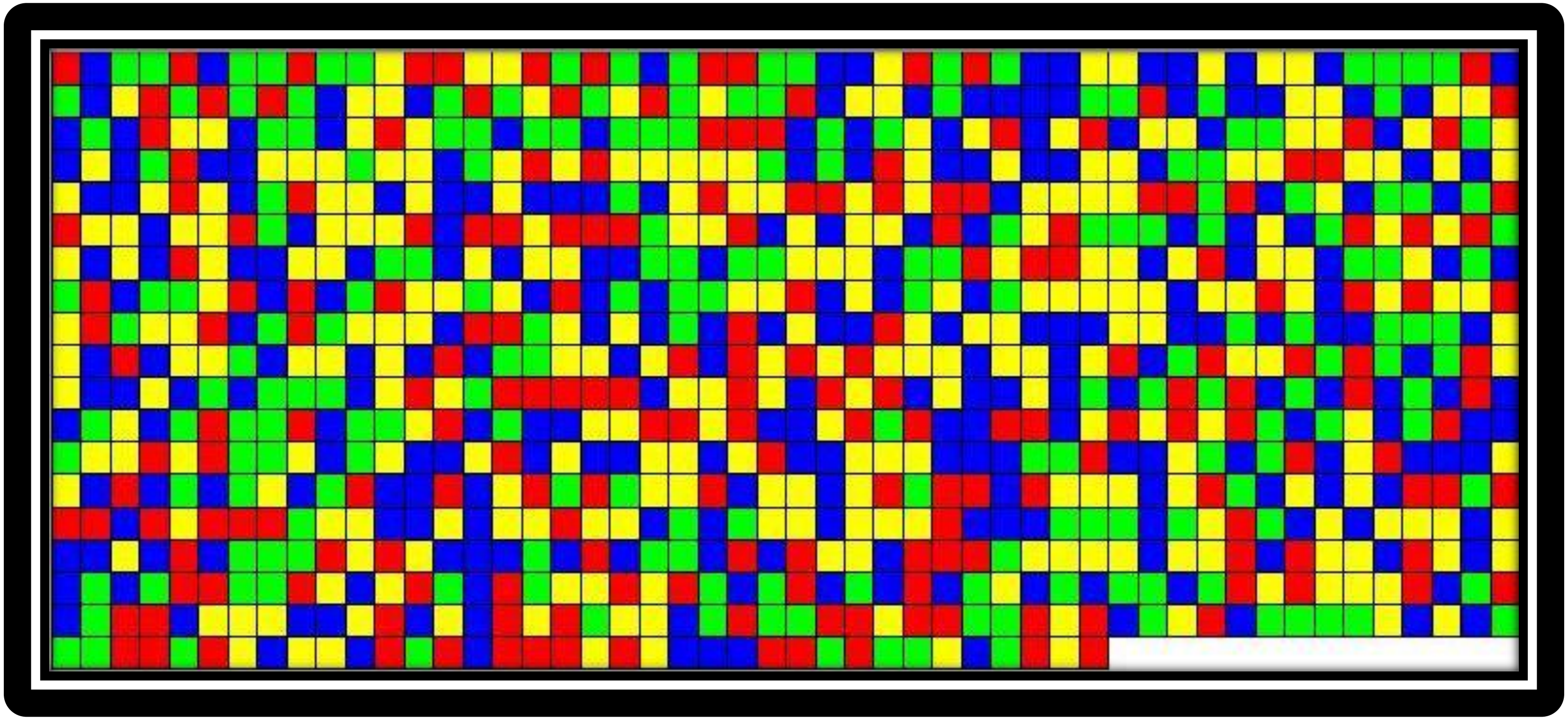
What it does

Let us visit an article….

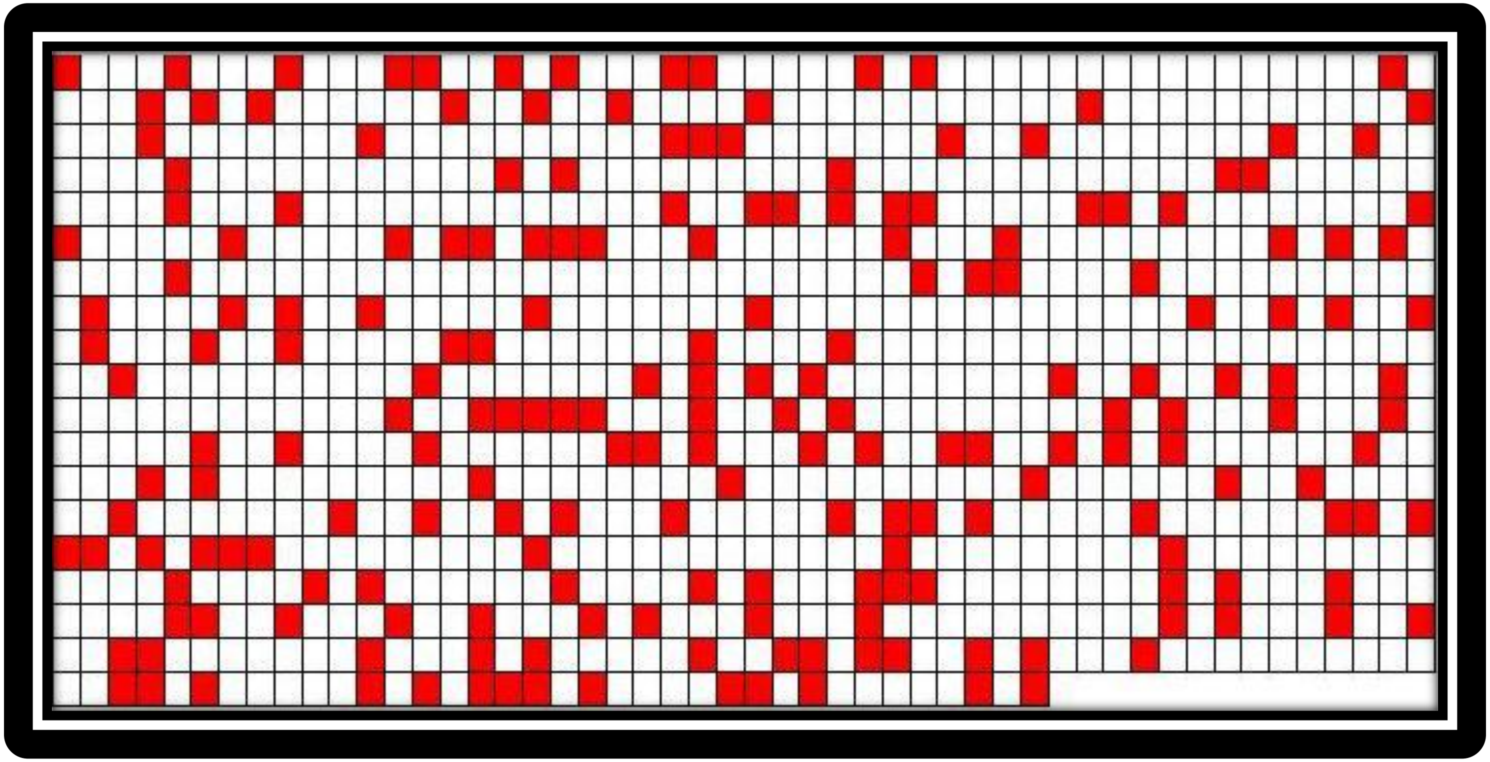
OUR IMMEDIATE RESEARCH ENDEAVORS

Fractals and Mathematical Morphology **in Deciphering The Quantitative Content in DNA sequences.**

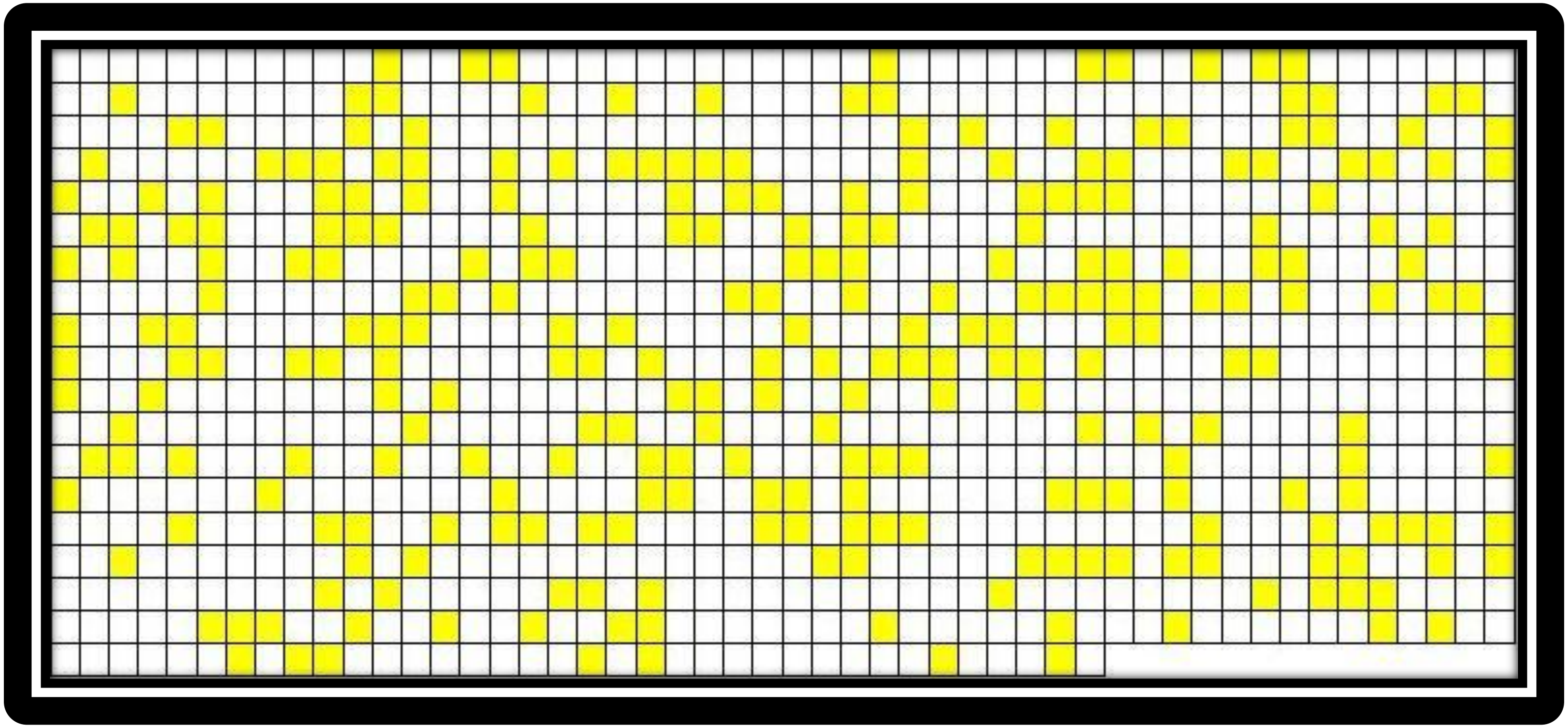
Color Template of DNA, OR1D2



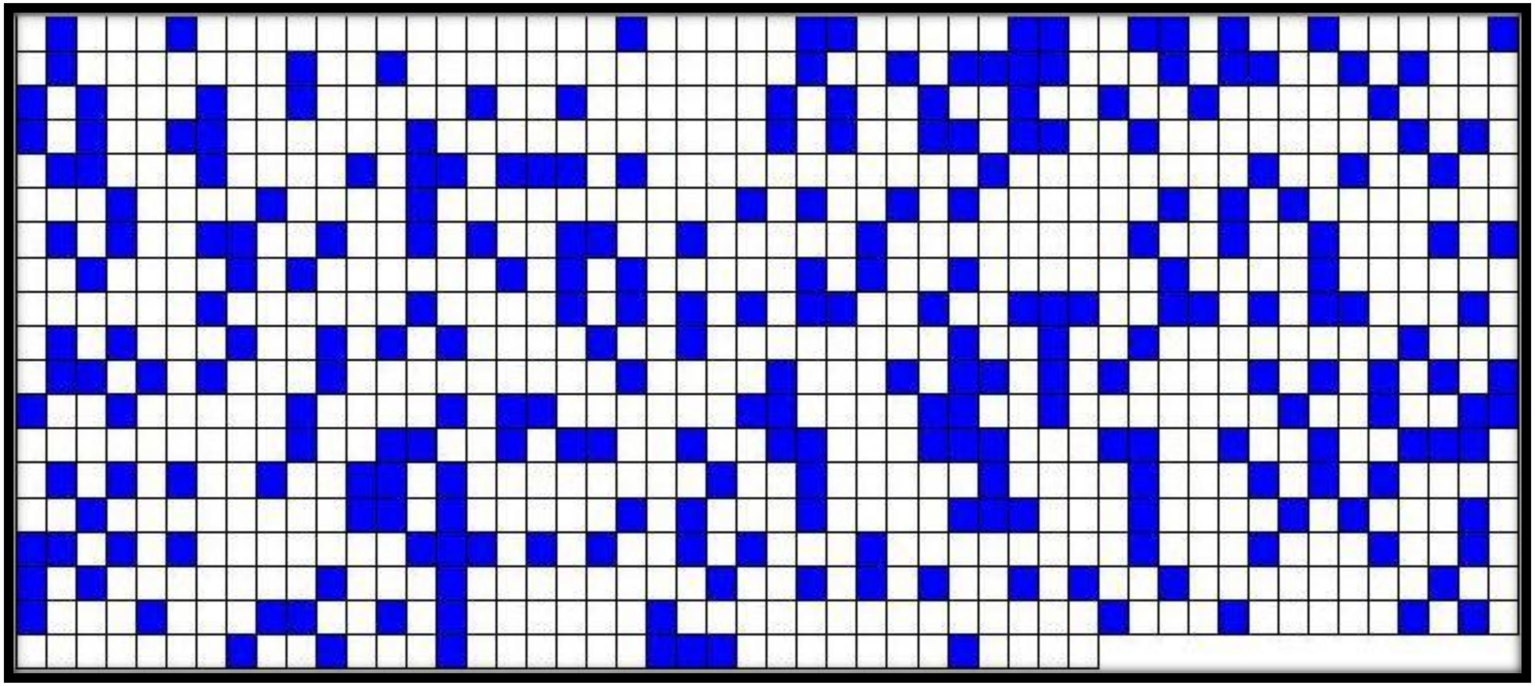
Color Template of ‘A’, OR1D2



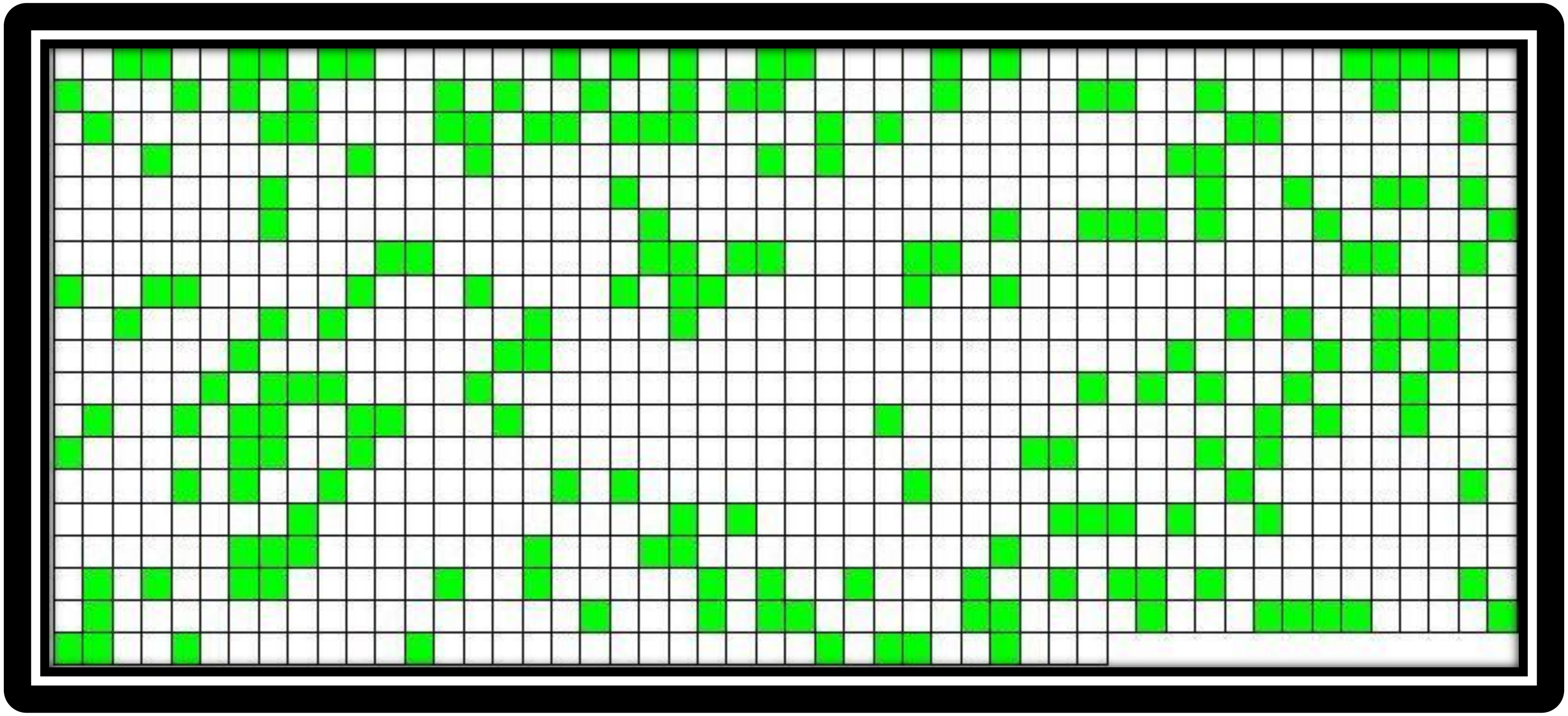
Color Template ‘T’, OR1D2



Color Template of ‘C’, OR1D2



Color Template ‘G’, OR1D2



		🟩	🟩			🟩	🟩		🟩	🟩							🟩		🟩		🟩			🟩	🟩					🟩		🟩												🟩	🟩	🟩	🟩		
🟩				🟩		🟩		🟩					🟩		🟩			🟩			🟩		🟩	🟩						🟩					🟩	🟩			🟩						🟩				
	🟩						🟩	🟩					🟩	🟩		🟩	🟩		🟩	🟩	🟩					🟩		🟩												🟩	🟩							🟩	
			🟩							🟩				🟩										🟩		🟩												🟩	🟩										
							🟩												🟩																				🟩			🟩			🟩	🟩		🟩	
							🟩													🟩												🟩			🟩	🟩	🟩		🟩				🟩						🟩
											🟩	🟩									🟩	🟩		🟩	🟩					🟩	🟩														🟩	🟩			🟩	
🟩			🟩	🟩						🟩				🟩					🟩		🟩	🟩							🟩			🟩																	
		🟩					🟩		🟩							🟩					🟩																			🟩		🟩			🟩	🟩	🟩		
						🟩									🟩	🟩																						🟩					🟩		🟩		🟩		
					🟩		🟩	🟩	🟩					🟩																					🟩		🟩		🟩			🟩				🟩			
	🟩			🟩		🟩	🟩			🟩	🟩				🟩													🟩													🟩		🟩			🟩			
🟩						🟩	🟩			🟩																							🟩	🟩					🟩		🟩								
				🟩		🟩			🟩								🟩		🟩										🟩										🟩								🟩	
								🟩													🟩		🟩											🟩	🟩	🟩		🟩			🟩								
						🟩	🟩	🟩								🟩				🟩	🟩											🟩																	
	🟩		🟩			🟩	🟩						🟩			🟩						🟩		🟩			🟩				🟩			🟩		🟩	🟩		🟩									🟩	
	🟩																	🟩				🟩		🟩	🟩						🟩	🟩					🟩				🟩	🟩	🟩	🟩					🟩
🟩	🟩			🟩								🟩														🟩		🟩	🟩			🟩																	

Different Parameters

- Poly string Mean
- Poly string SD
- Fractal Dimension
- Hurst Exponent
- Succolarity

In OR1D2

- The mean of A is 1.200000
- The SD of A is 0.537970
- The mean of G is 1.333333
- The SD of G is 0.620876
- The mean of T is 1.210280
- The SD of T is 0.509431
- The mean of C is 1.465608
- The SD of C is 0.745450

Succolarity: A Measure for Continuous Density

- Hurst Exponent- 0.598911
- Fractal Dimension: 1.401089

Succolarity of The Template of A: 0.001026
Succolarity of The Template of T: 0.001690
Succolarity of The Template of G: 0.000522
Succolarity of The Template of C: 0.001482

The resultant fact

- We have enumerated all such parameters for all the Human Ors. We are about to design a database consisting all the results and classification based on these parameters.
- Now any foreign sequence(generated by an L-System) could be used to classify in the class of Ors based on the above parameters. This is basically hitting a specific member or members. That is, we could make another receptor following some rule and that is best fitted.

References

- SK. S. Hassan, Pabitra Pal Choudhury, R. L. Brahmachary, Arunava Goswami, **Designing exons for human olfactory receptor gene subfamilies using a mathematical paradigm , Journal of Bioscience, Springer, Vol 35 (3), September, 2010.**
- SK. S. Hassan, Pabitra Pal Choudhury, R. L. Brahmachary, Arunava Goswami, **L-Systems: A Mathematical Paradigm For Designing Full Length Genes And Genomes**, ***GJCST (2010) 10: 119-122, USA.***
- Sk. Sarif Hassan, Pabitra Pal Choudhury, Amita Pal, R. L. Brahmachary, Arunava Goswami **‘Combination of L-systems: For Designing Human Olfactory Receptor Pseudo-gene, OR1D3P (In press of IJCC, USA, August, 2010)**

Further References…

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

- Young J-M, Endicott R-M, Parghi S-S, Walker M, Kidd J-M, Trask B-J (**2008**) **Extensive copy number variation of the human olfactory receptor gene family.** **Am. J. Hum. Genet. 83: 228–242.**
- Malnic B, Hirono J, Sato T, Buck L-B (**1999**) **Combinatorial receptor codes for odors.** **Cell 96: 713-723.**
- Glusman G, Yanai I, Rubin I, Lancet D (**2001**) **The complete human olfactory subgenome.** **Genome Research 11: 685-702.**
- Prusinkiewicz P, Lindenmayer A (**1990**) in **The algorithmic beauty of plants,** **Springer-Verlag ISBN 978-0387972978.**
- Yu Zu-Guo et al **2002** **Fractals in DNA sequence analysis** ,Chinese Phys. 11 1313-1318
- Claire Ainsworth et al , **September, 2009**, **Cells go fractal**: Nature
- B. Mandelbrot., The fractal Geometry of Nature, **1975**.
- Bransley, Fractals Everywhere. Springer Verlag.

My heartiest thanks to...

- **Dr. Arunava Goswami, AERU, ISI**
- **Prof. Pabitra Pal Choudhury, ASU, ISI,**
- **Dr. Amita Pal, BIRU, ISI**
- **Prof. R. L. Brahmachary, Ex Professor, ISI.**

Thank You