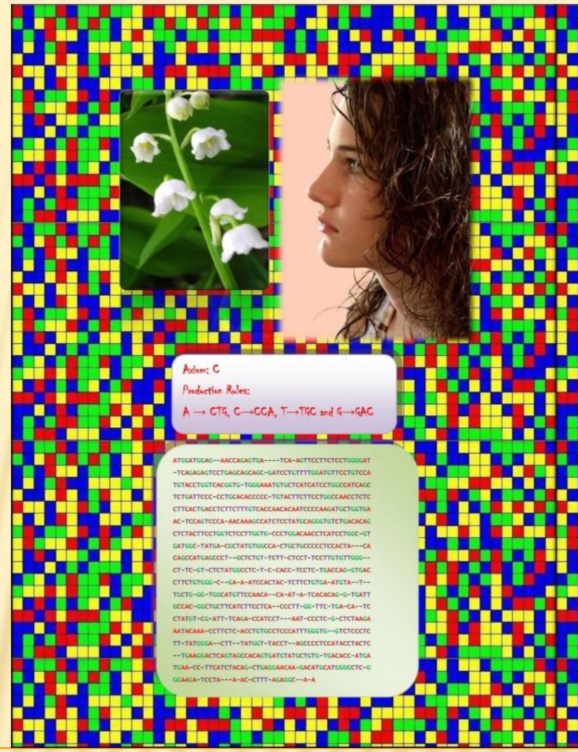
L-SYSTEMS: A MATHEMATICAL PARADIGM FOR DESIGNING GENES AND GENOME WITH MORPHOLOGICAL FLAVOR



Pabitra Pal Choudhury

Applied Statistics Unit

Indian Statistical Institute

Kolkata

ISI Bangalore

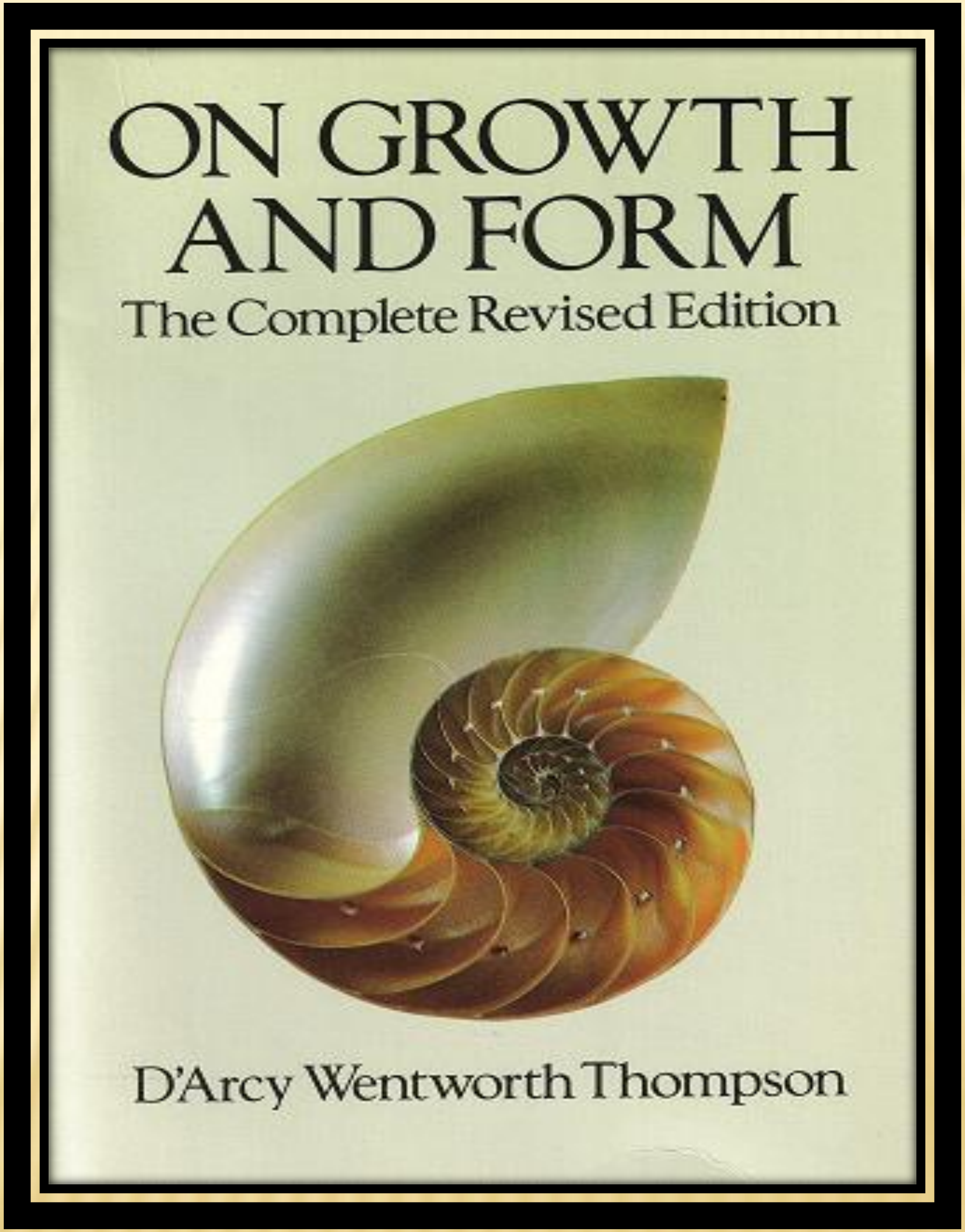
26th October, 2010

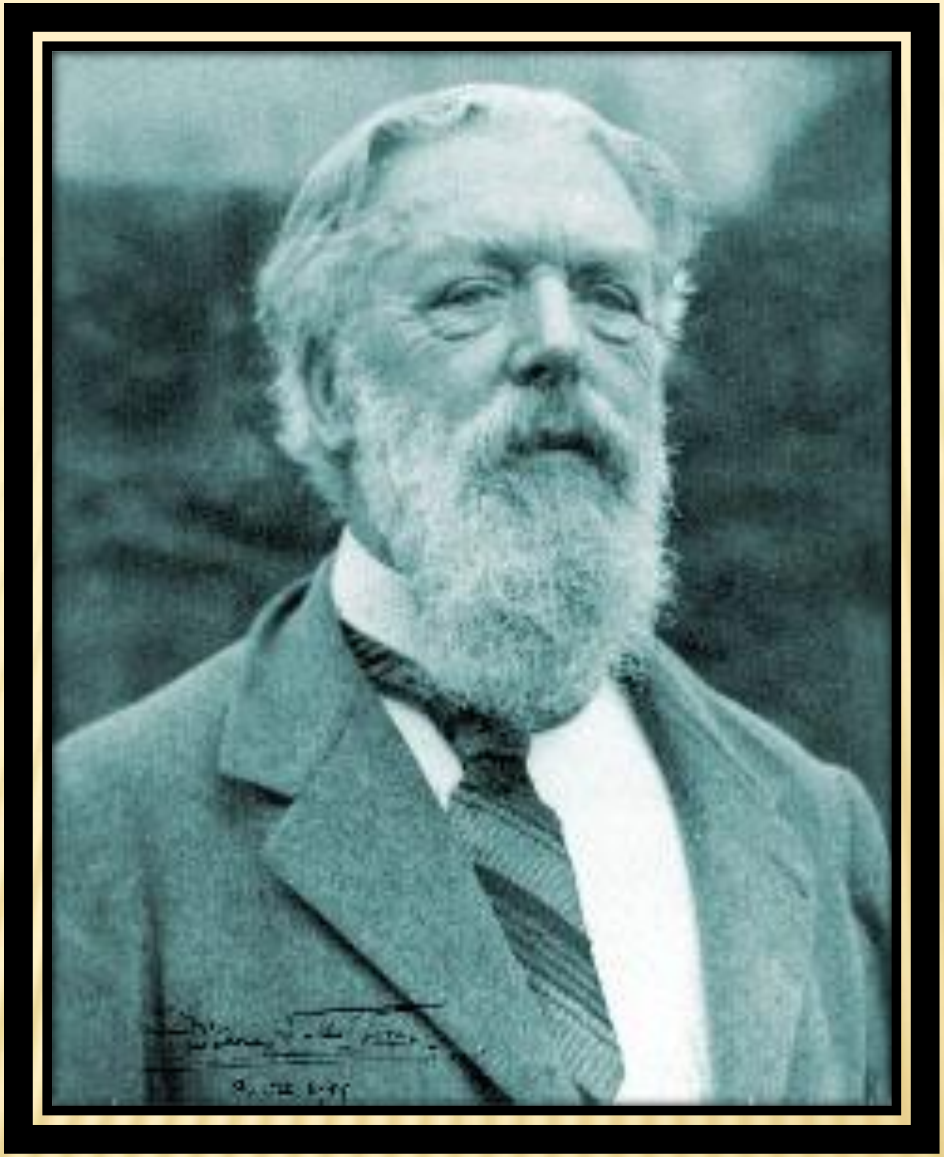
The beauty of symmetric (and asymmetric) pattern in the living world, as in flower or a nautilus shell has been noticed by people since the earliest times.

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 inking 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.

Put spiral equation

Nautilus Shell

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.

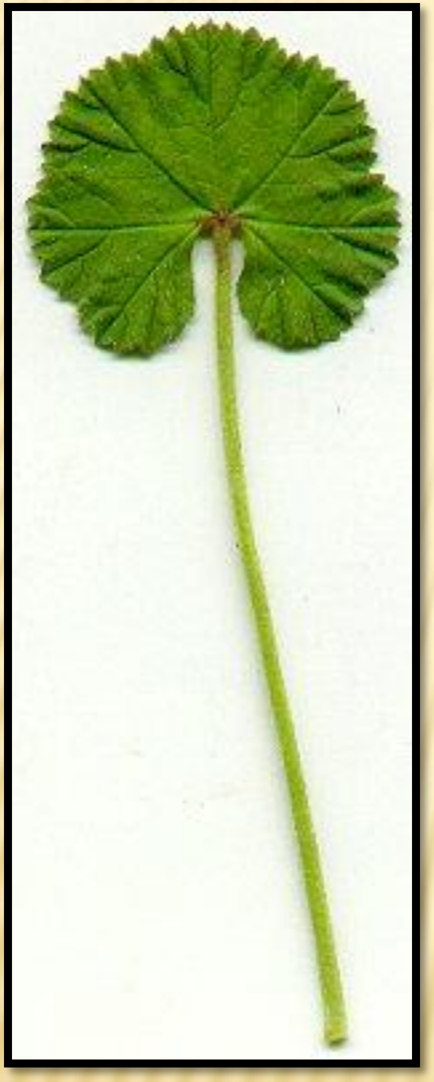
$$r = ae^{b\theta} \text{ where } a, b \text{ are parameters.}$$

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

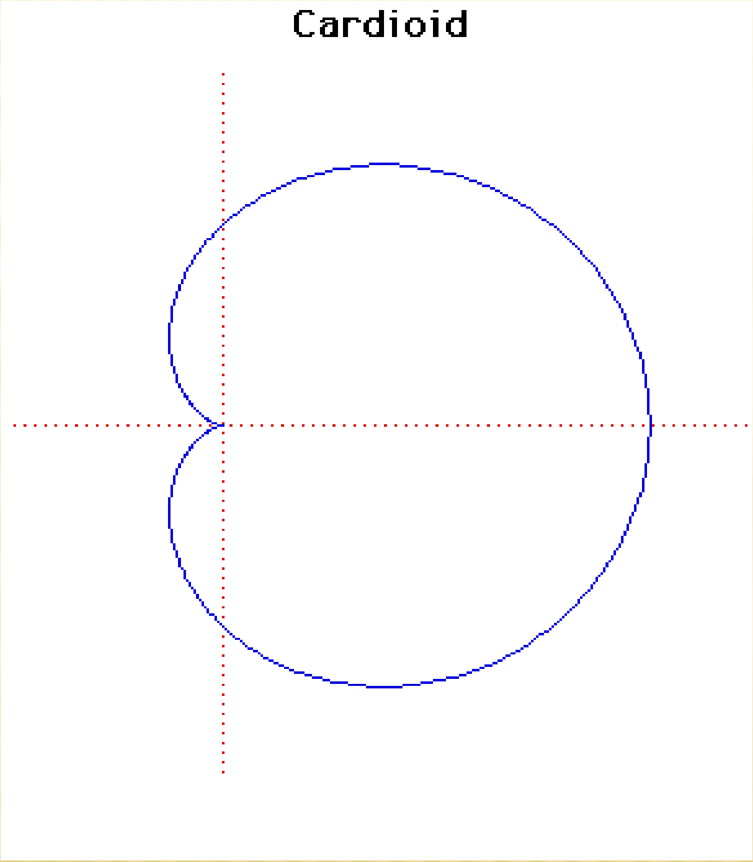
Likewise, Thompson described a curve resembling the outline of a reniform leaf. Variations of the formula r=sin(ϕ/2) can generate various shapes.

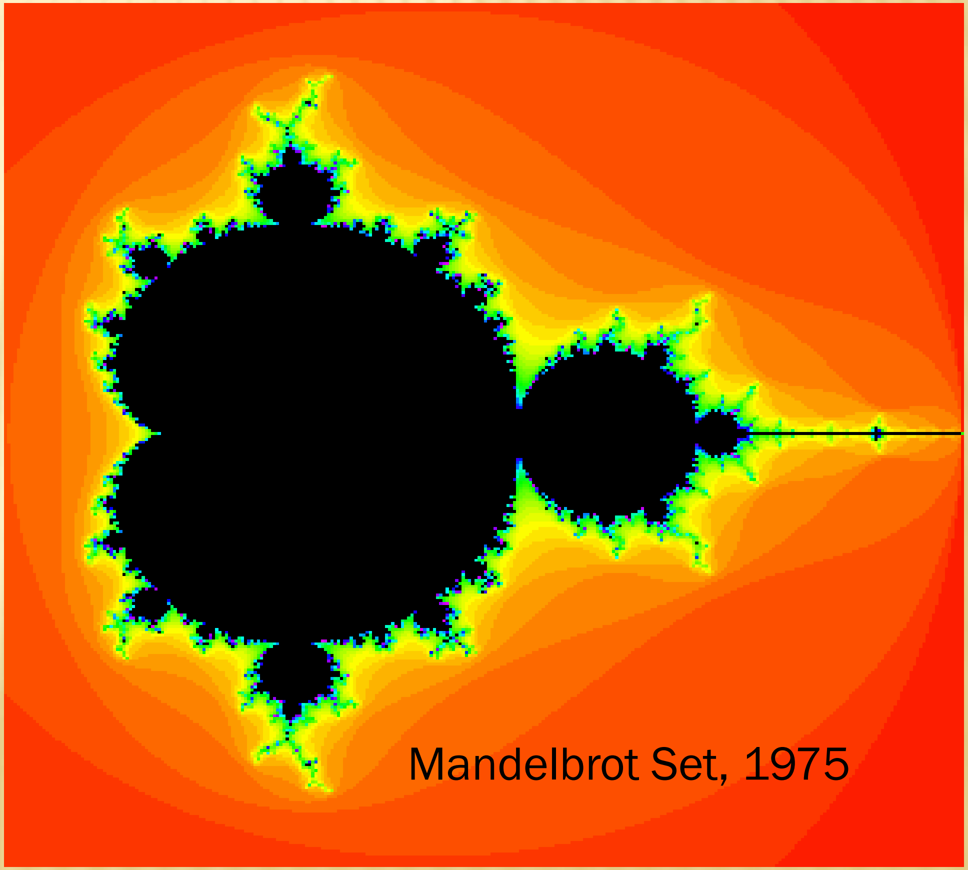
The late T. A. Davis, ISI, worked on Sunflower head.

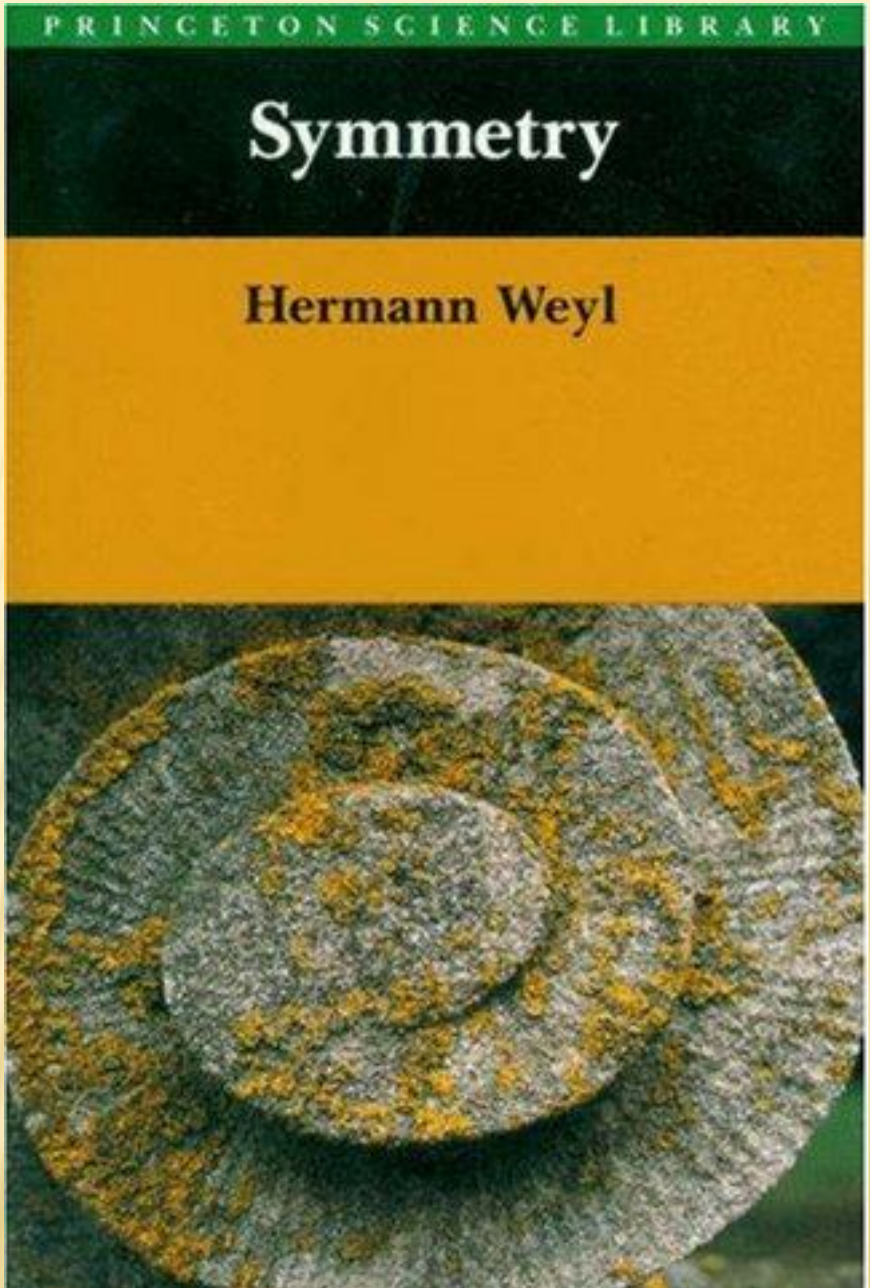


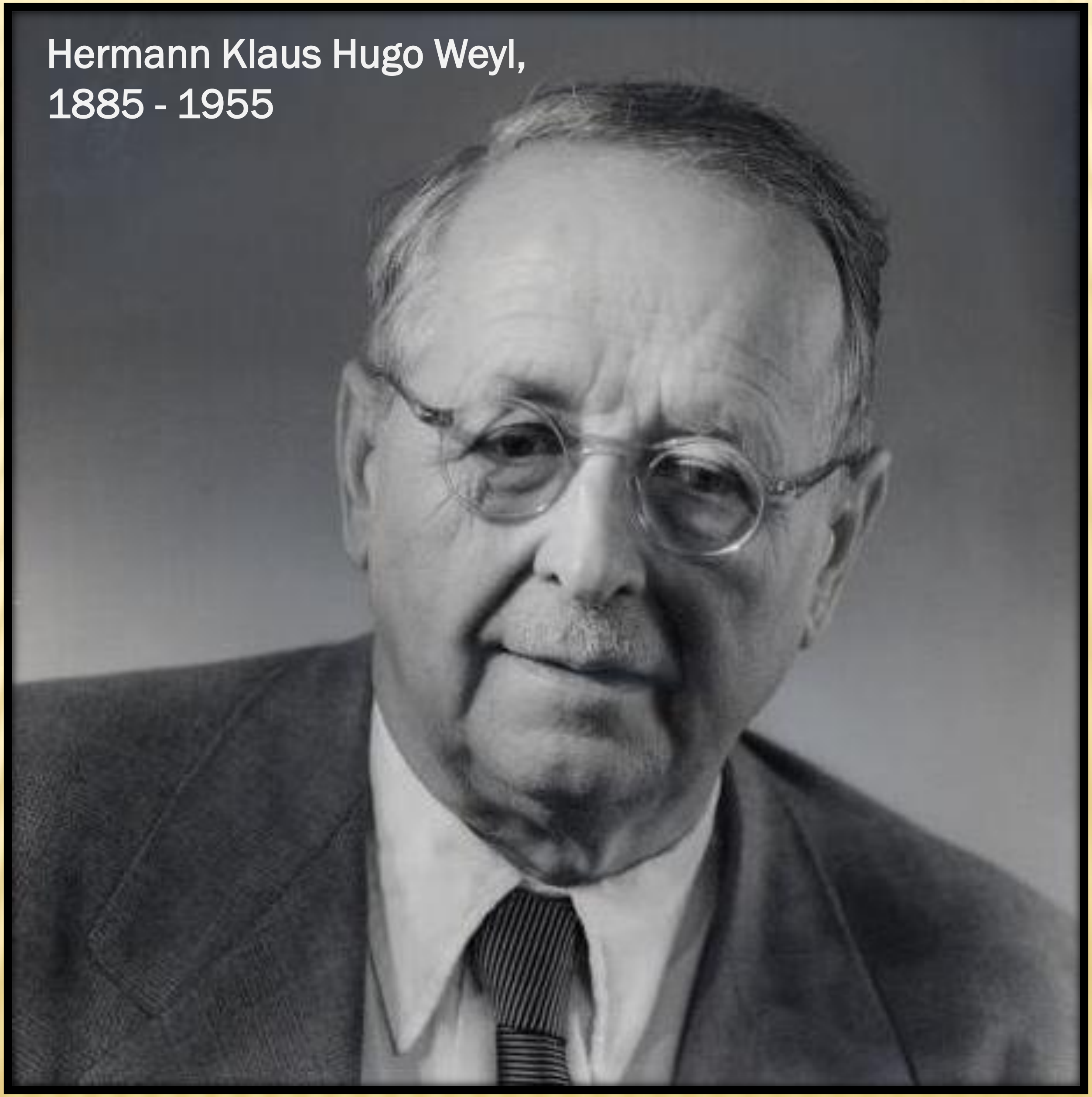
The symmetry pattern in the spirals of the sunflower head was noticed by Thompson. Around 1970, T. A. Davis of the Indian Statistical Institute worked on this aspect. Some details are sketched below:

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

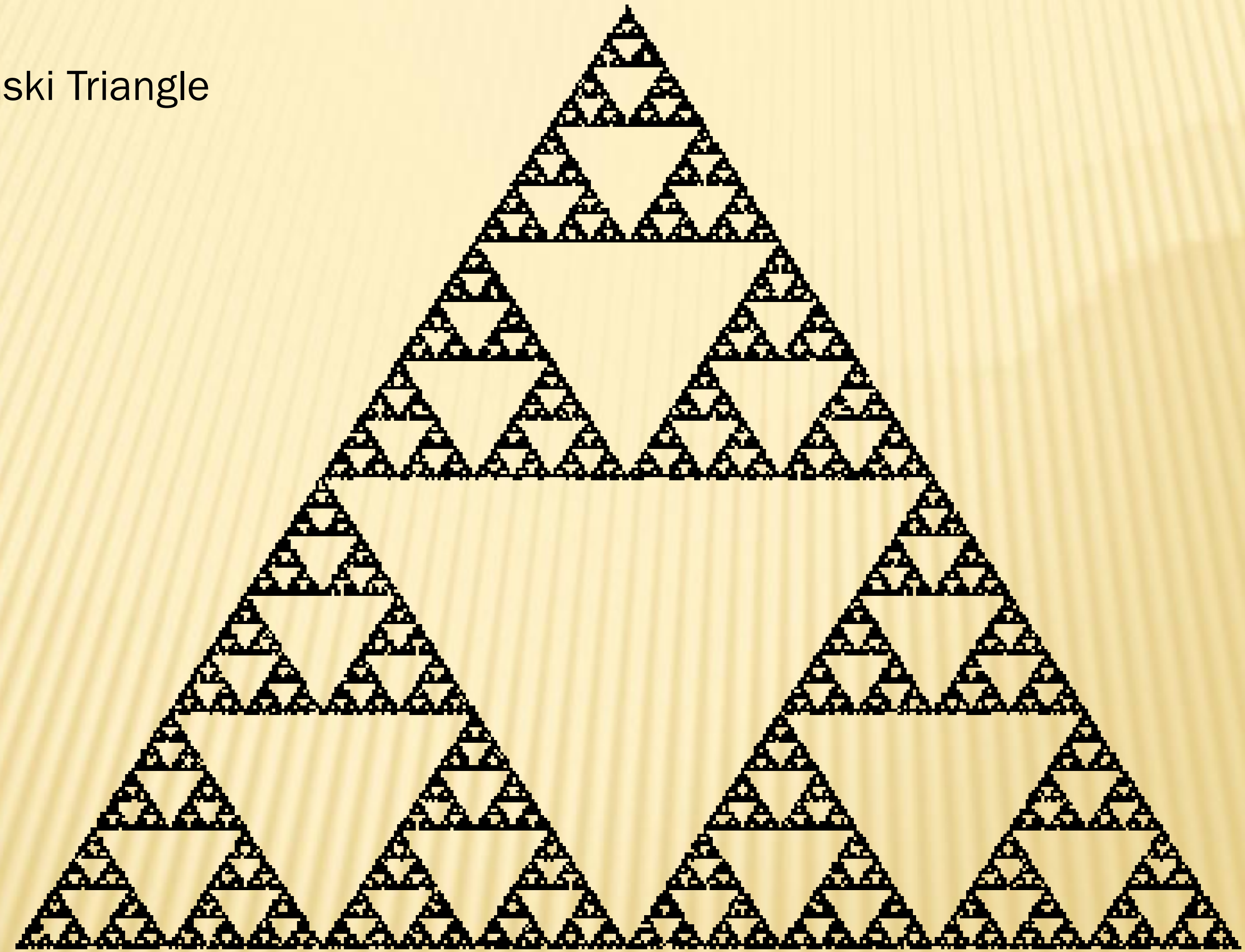
In 1951, Hermann Klaus Hugo Weyl, a great early exponent of Relativity theory and, Group Theory & Quantum mechanics and also wrote a book, SYMMETRY.

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

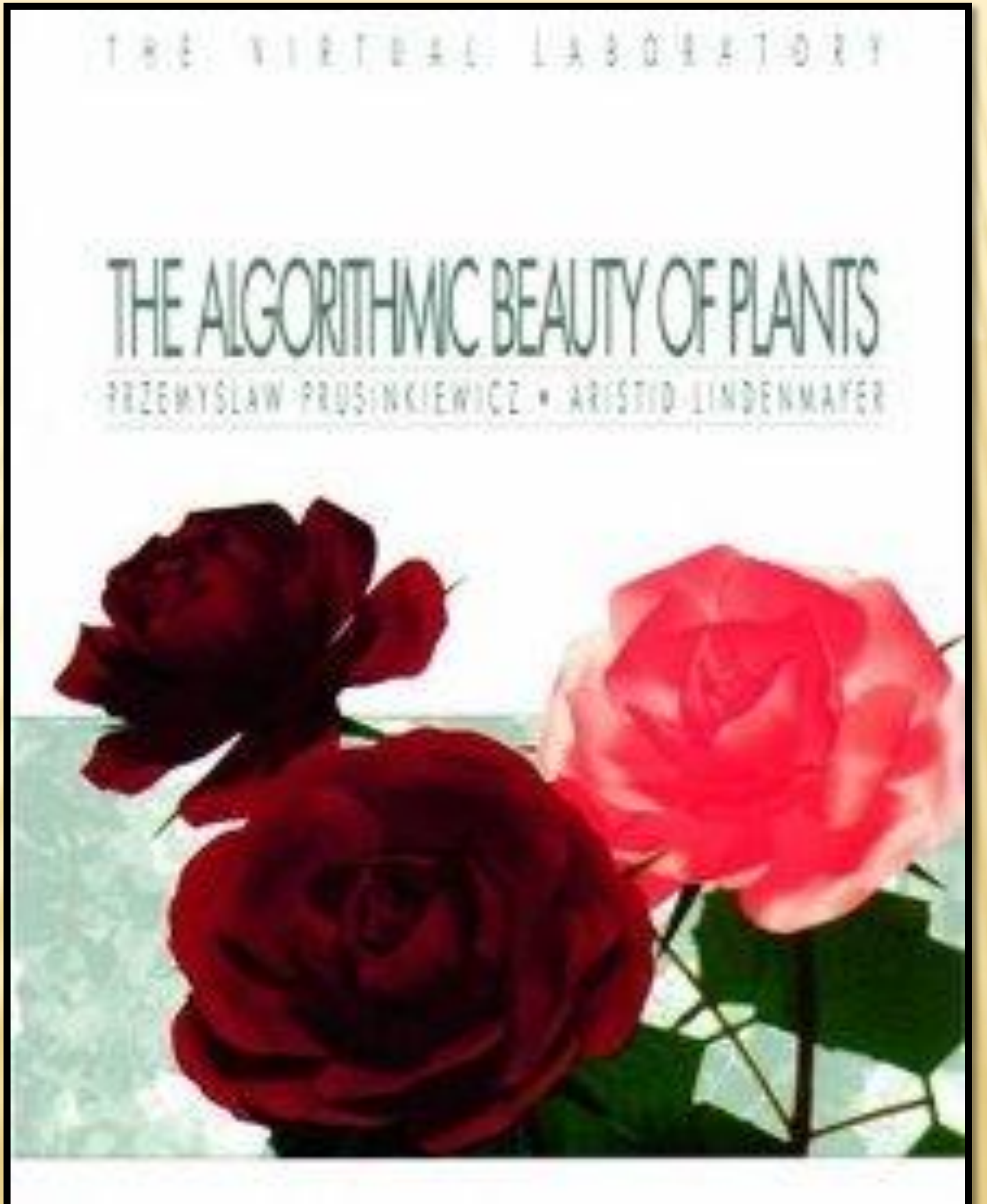
Sierpinski Triangle



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

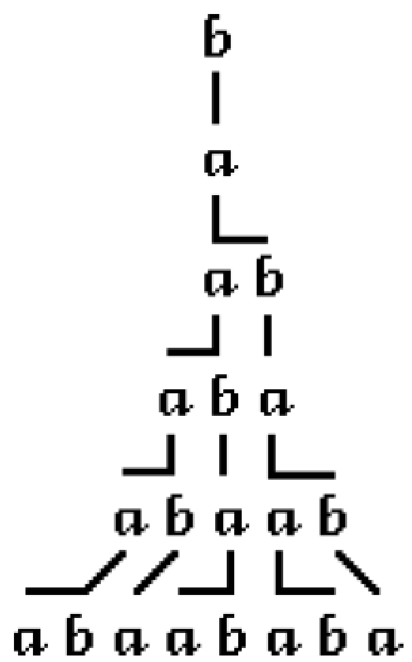
Aristid Lindenmayer (November 17, 1925 – October 30, 1989) was a Hungarian biologist

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:

$$a \rightarrow ab$$
$$b \rightarrow a$$



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.

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

- OR denotes Olfactory Receptors.
- 1 denotes family name.
- D denotes subfamily name.
- 2 denotes the member name.

WHY OR1D2?

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

- One of them, OR1D2, binds to Bourgeonal, a volatile chemical constituent of the fragrance of **Lily of the valley** or **Our Lady's tears**, **Convallaria majalis** (also the **national flower of Finland). Picture of the flower should be attached**

>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/

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, this may not be 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?

This observation motivated us…

OR1D2, OR1D4 AND OR1D5

- **<u>OR1D2</u> was used as query in BLASTn search for similar ORs in HORDE and <u>OR1D4</u> and <u>OR1D5</u> 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.**
- **As per HORDE, ORs with >80% nucleotide match are called members of the same subfamily of OR. >60% : OR Family.**

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

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 we could be able to find out the selection key (the functioning part for the olfaction purpose).**

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 created DNA sequence is at all functioning or not functioning.**

A NEW PRINCIPLE

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…

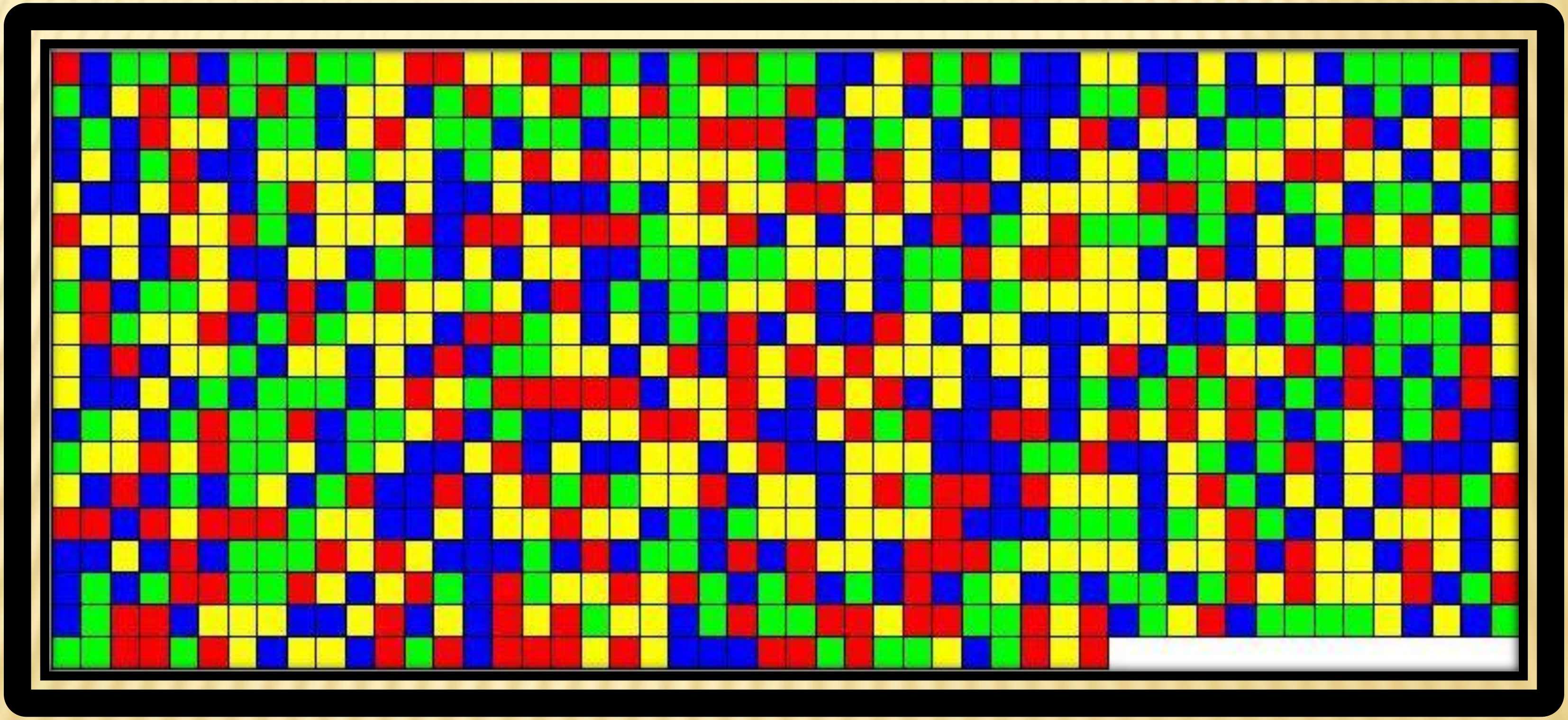
BUT, IS THE PROCEDURE HAVE SOME NOVELTY?

The beauty of the proposed methodology would enlighten us to answers of some unanswered questions in "Evolutionary Science".

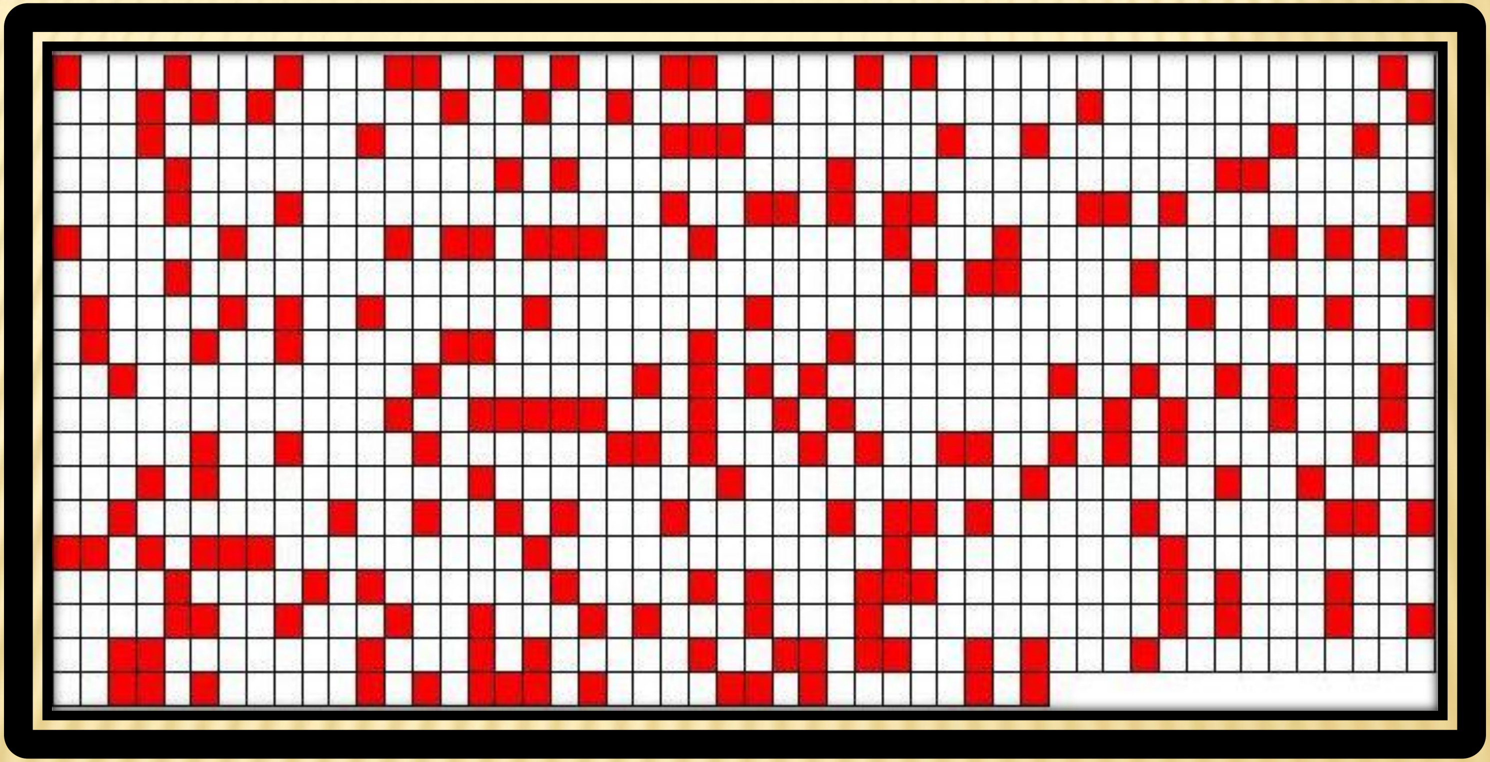
OUR IMMEDIATE RESEARCH ENDEAVORS

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

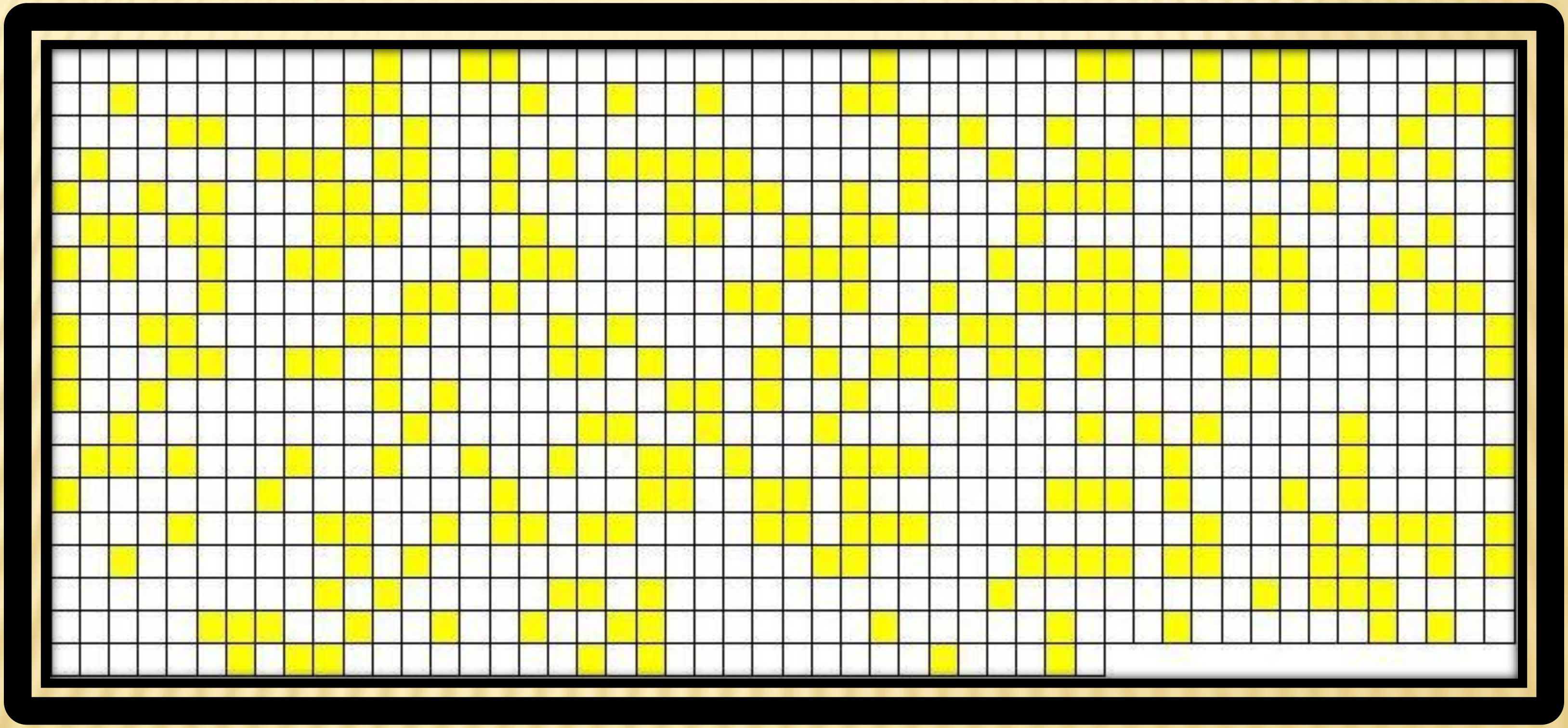
COLOR TEMPLATE OF DNA, OR1D2



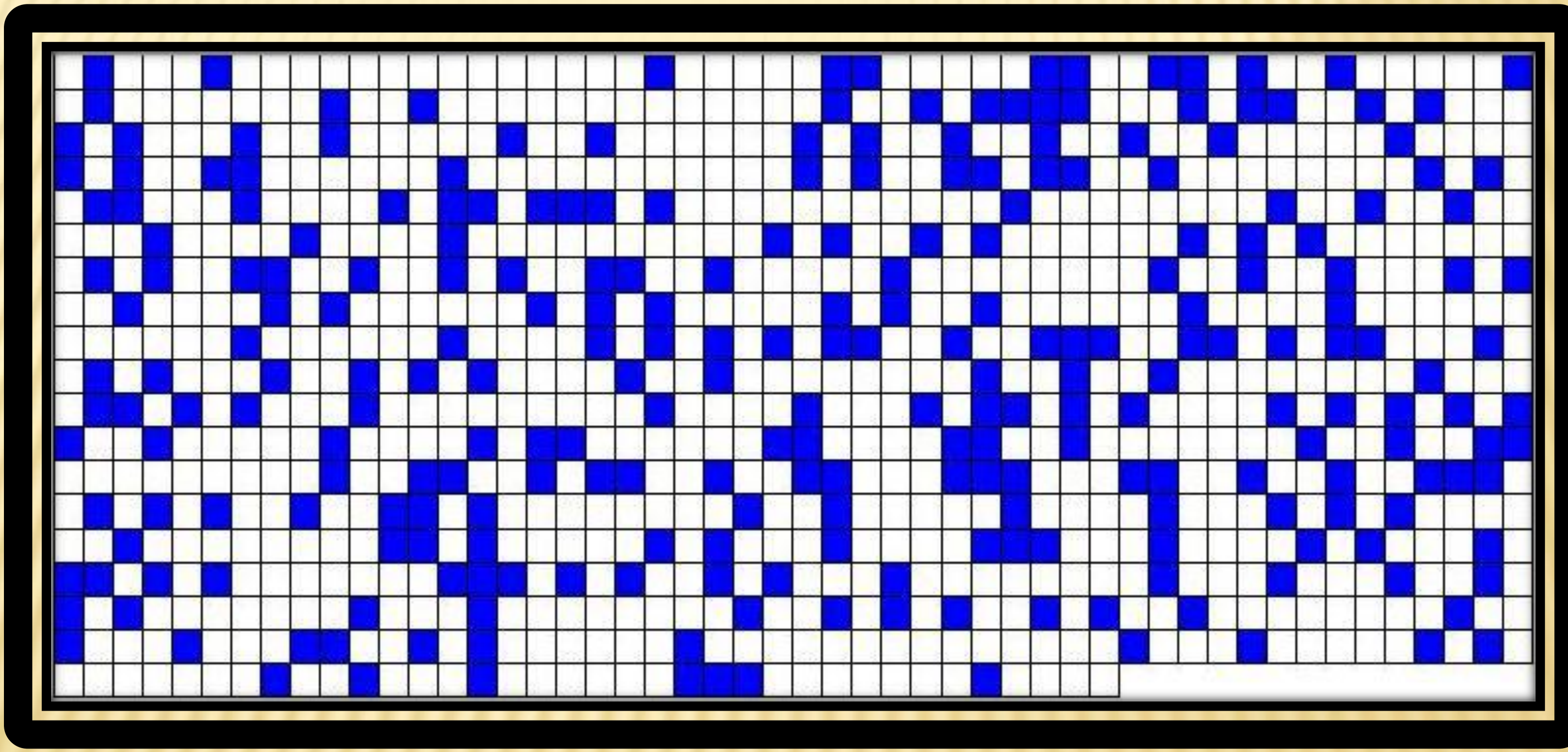
COLOR TEMPLATE OF 'A', OR1D2



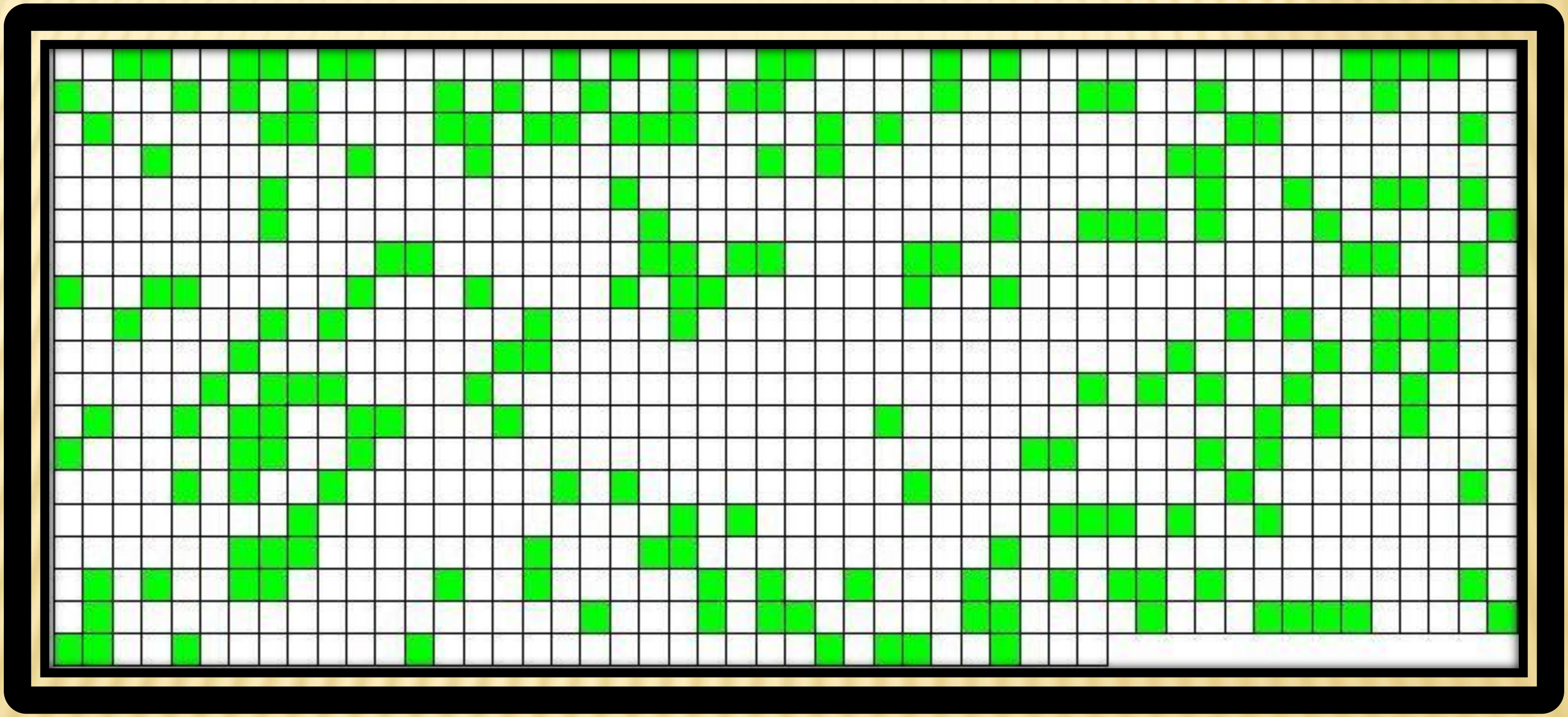
COLOR TEMPLATE 'T', OR1D2



COLOR TEMPLATE OF ‘C’, OR1D2



COLOR TEMPLATE ‘G’, OR1D2



SUCCOLARITY: A MEASURE FOR CONTINUOUS DENSITY

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

DNA SEQUENCE: A MULTI-FRACTAL

The Texture of A, T, C, and G generate Fractals.

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.**
- SK. S. Hassan, Pabitra Pal Choudhury, Arunava Goswami, R. L. Brahmachary **Designing exons for human olfactory receptor gene subfamilies using a mathematical paradigm** under editorial board review in **Proc. Natl. Acad. Sci U. S. A (October, 2009).**
- 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.**

REFERENCES…

- 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.
- **P. P. Choudhury, S. Sahoo, B. K Nayak, and Sk. S. Hassan**, Carry Value Transformation: It's Application in Fractal Formation 2009 IEEE International Advanced Computing Conference (IACC 2009), Patiala, India, 6-7 March, pp 2613-2618, 2009

Questions Please!


2classnotes
Thanking You