 doi:10.20944/preprints202103.0450.v2

Physical Laws shape up HOX Gene Collinearity

Spyros Papageorgiou
Institute of Biosciences and Applications, National Center for Scientific Research 'Demokritos'
153 10 Athens, Greece
Email: spapage@bio.demokritos.gr

Abstract

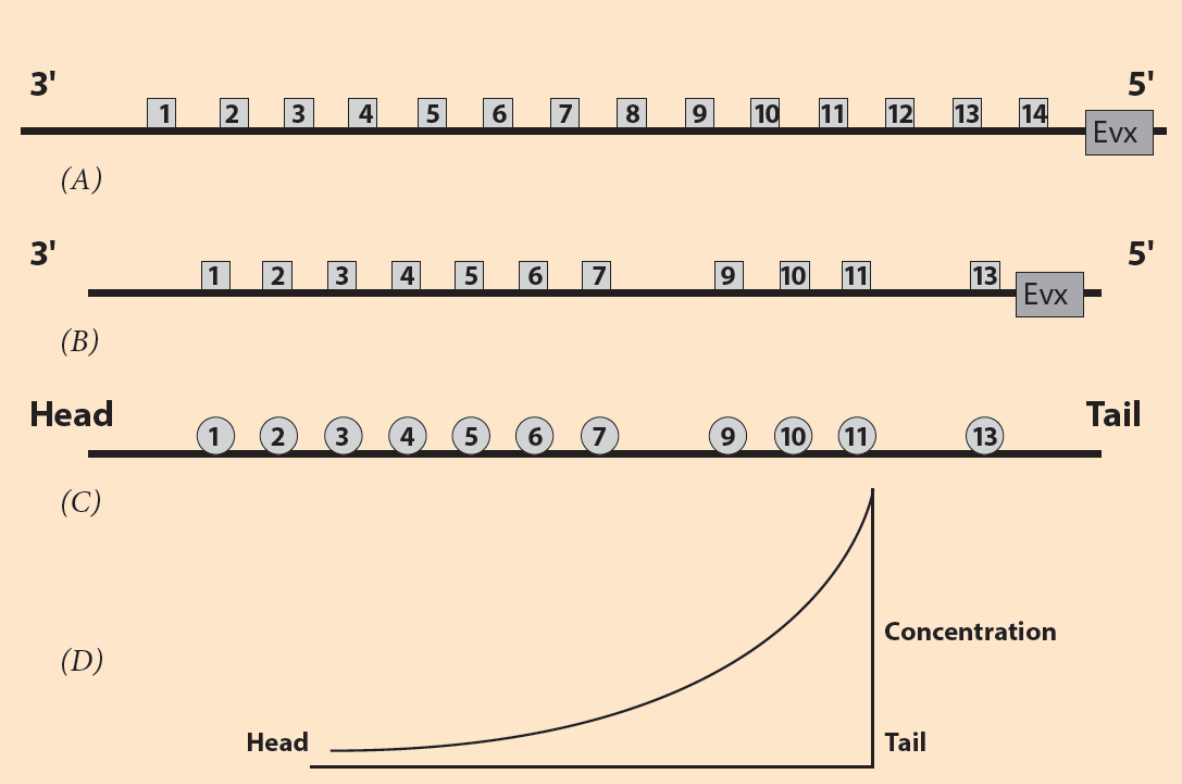
Hox gene collinearity (HGC) is a multiscalar property of many animal phyla particularly important in embryogenesis. It relates entities and events occurring in Hox clusters inside the chromosome DNA and in embryonic tissues. These two entities differ in linear size by more than four orders of magnitude. HGC is observed as spatial collinearity (SC) where the Hox genes are located in the order (Hox1, Hox2, Hox3 …) along the 3' to 5' direction of DNA in the genome and a corresponding sequence of ontogenetic units (E1, E2, E3, …) located along the Anterior – Posterior axis of the embyo. Expression of Hox1 occurs in E1. Hox2 in E2, Hox3 in E3… Besides SC, a temporal collinearity (TC) has been also observed in many vertebrates. According to TC first is Hox1 expressed in E1, later is Hox2 expressed in E2, followed by Hox3 in E3,… Lately doubt has been raised whether TC really exists. A biophysical model (BM) was formulated and tested during the last twenty years. According to BM, physical forces are created which pull the Hox genes one after the other driving them to a transcription factory domain where they are transcribed. The existing experiments support this BM description. Symmetry is a physical-mathematical property of Matter that was explored in depth by Noether who formulated a ground-breaking theory that applies to all sizes of Matter. This theory applied to Biology can explain the origin of HGC as applied not only to animals developing along the A/P axis but also to animals with circular symmetry.

Keywords: Hox gene collinearity, spatial temporal collinearity, vertebrates, Noether theory.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

1 Introduction

Hox Gene Collinearity (HGC) is a basic embryonic property coordinating development of vertebrates and many other animal phyla. It was first observed by E.B. Lewis in the *Drosophila* BX-C gene complex [1]. Lewis noticed that a class of Hox genes are located in clusters following an ordered sequence (Hox1, Hox2, Hox3,…) along the direction 3' to 5' on the chromosome (Fig. 1). These genes are expressed in the same order in the embryo along the Anterior/Posterior (A/P) axis. This common order in the chromosome and the embryo is denoted spatial collinearity (SC). SC is a surprising property because it correlates spatial entities differing by about 4 orders of magnitude. The linear dimension of an early vertebrate embryo is about 1mm, whereas the linear dimension of a Hox cluster is about 100 nm [2]. Biomolecular mechanisms by themselves alone cannot explain such correlations.



[Figure 1]

Ordering of Hox genes and the sequence of the ontogenetic units. (A) Hox gene ordering of a (theoretical) common ancestor. Gene **Evx** is located next to the 5'end of the Hox cluster. (B) Ordering of the mouse HoxA cluster. Hox8 and Hox12 are missing. (C) The corresponding ontogenetic units of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

mouse along the A/P axis. (D) The steady state monotonic concentration gradient of a morphogen. The peak is at the tail region.

SC is not the only property controlling vertebrate embryonic growth. Another principle for vertebrate development was later established: temporal collinearity (TC) [3, 4]. According to TC, the time dependence of Hox gene expressions follows the empirical rule: in the ordered sequence of Hox genes from the telomeric to the centromeric end of the Hox cluster, there is a corresponding sequence of embryonic ontogenetic units E1, E2,… along the anterior-posterior axis. First is Hox1 expressed in E1 followed later by Hox2 expressed in E2, etc.[3, 4].

In the evolutionary process, a Hox cluster may appear in several homologues as a result of whole genome duplications (WGD) [5]. For instance, in vertebrates there are four homologue Hox clusters denoted HoxA, HoxB, HoxC, HoxD. These homologue clusters cooperate for the normal embryonic development. The ordered Hox genes of a cluster (Hox1, Hox2 ,..) constitute a paralogy group Pg [4, 5]. For instance Pg1 can be traced in Hox clusters of different animal as a result of their origin from a common ancestor Hox1. In vertebrates the paralogy group consists of 13 paralogue Hox genes Pg1, Pg2…,Pg13 whose role is crucial (see Section 4). In many cases, due to whole genome duplications, the genomes contain several copies of the Hox clusters. In a comprehensive study of Hox gene expressions in *Xenopus laevis*, M. Kondo et al. analyzed the WGD and in particular the allotetraploidization which they estimated that it occurred 17 -18 Mya [5]. They designated the two subgenomes ‘homologs’ L and S and compared the homolog gene expressions within a Hox gene cluster. They concluded that many L, S homologs are not orderly correlated with the paralogue order as would be expected according to TC. They suggested that ‘the TC hypothesis must be revisited by comprehensive analysis of the developmental timing of transcriptional initiation of Hox genes…’ [5] (See Section 3).

From a different point of view, the above doubt of TC validity was challenged by Durston et al. indicating that TC is verified in many vertebrates including cephalochordates [6, 7]. In particular, TC in cooperation with time-space transformations leads to the development of different body parts of the vertebrate embryo [7].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

In view of the above ambiguity, it is worth interpreting the recent contradicting data applying a Biophysical Model (BM) which provides a unifying approach.

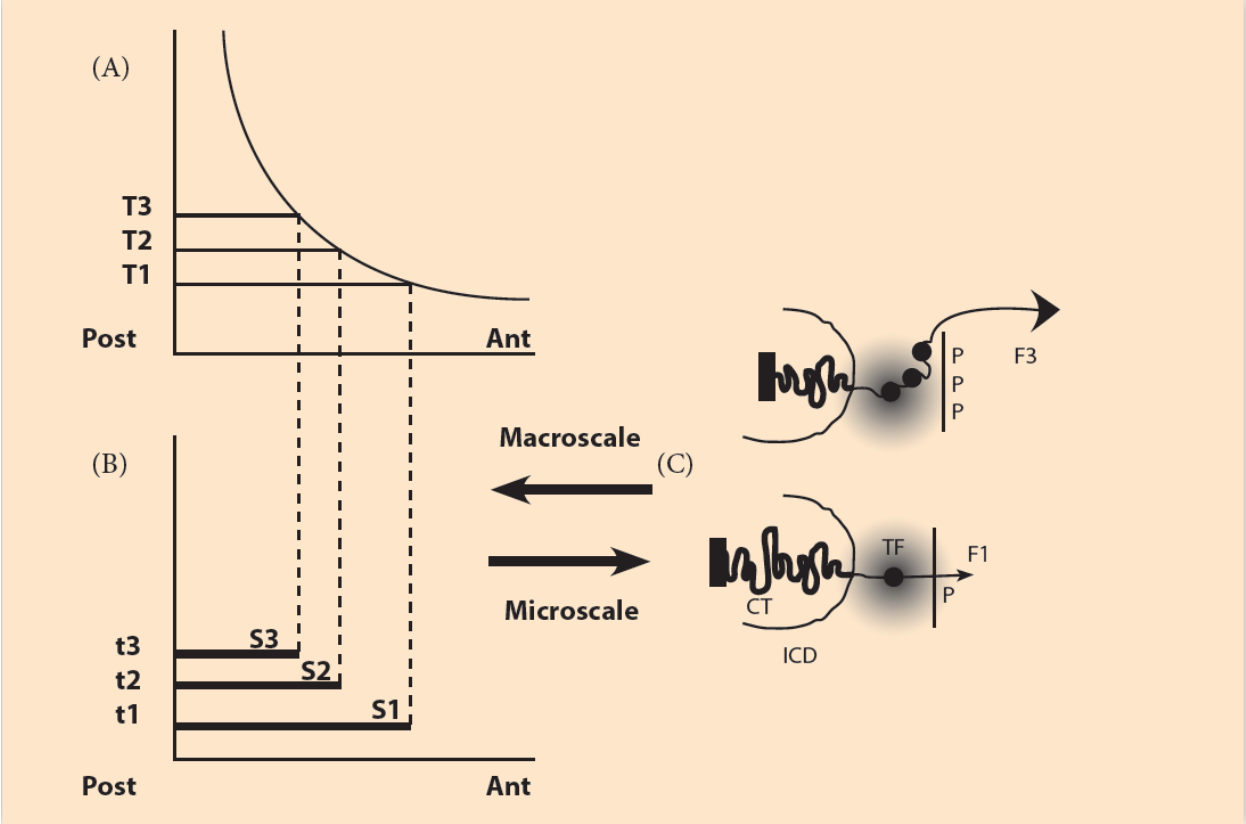
2 The Biophysical Model for Hox Gene activation

The multiscale nature of HGC motivated the formulation of the following Biophysical Model (BM). The basic hypothesis is that the pulling forces act on the Hox clusters and they are influenced from contributions from both scales. At the embryonic (macroscopic) scale the contribution is contained in a morphogen gradient along the Anterior-Posterior embryonic axis [8-10]. The contribution from the microscopic scale originates from the cluster itself. Before activation the Hox cluster is packed inside the chromatin territory (CT) in a compact unity (Fig, 2). No forces are created at this stage. Gene activation starts when attractive forces are gradually created. Such forces emerge when polar molecules (positively charged) are transferred and allocated at the telomeric region of the cluster. (Detailed experimental evidence for these events is found in [11]). The forces gradually increase pulling the genes out of the CT (Fig.1). Tentatively, the nature of the forces is electric [8,9].

A simple heuristic form for the attractive forces **F** acting on Hox clusters is the following [10,11]

$$\mathbf{F} = \mathbf{N} \times \mathbf{P} \qquad (1)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2



[Figure 2]

The macroscale morphogen gradient and the microscale Hox gene clustering in space and time (adapted from Papageorgiou S. [Biology 2017: 6, 32]). A) Concentration thresholds (T1,T2,T3) divide A/P axis in partially overlapping expression domains. B) The time sequence (t1, t2, t3) combined with thresholds sequence (T1, T2, T3) determine the Hox1, Hox2, Hox3 activation in space and time. S1, S2, S3 are the partially overlapping and nested expression domains of Hox1, Hox2, Hox3. C) (bottom) In an anterior cell of S1, a small force **F1** pulls Hox1 (black spot) out of CT toward the Interchromosome Domain (ICD) in the regime of the Transcription Factory (TF) (grey domain). Apposition of polar molecules P opposite the telomeric end of the Hox cluster (top). At a later stage in a more posterior location of S3, a stronger force **F3** pulls Hox1, Hox2, Hox3 out of CT in TF. Apposition of PPP molecules .

In eq. (1) factor **N** represents the contribution of the Hox cluster which is negatively charged in agreement with the overall negative charge of DNA. The 'positive' factor **P** is graded (low anteriorily- high posteriorily) and represents the embryonic contribution to **F**. The force **F** is an electric quasi- Coulomb force and applies at the telomeric end of the Hox cluster (Fig. 2). In BM the attractive force **F** provides an interplay between the microscopic scale of the Hox cluster and the macroscopic embryonic scale (Fig. 2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

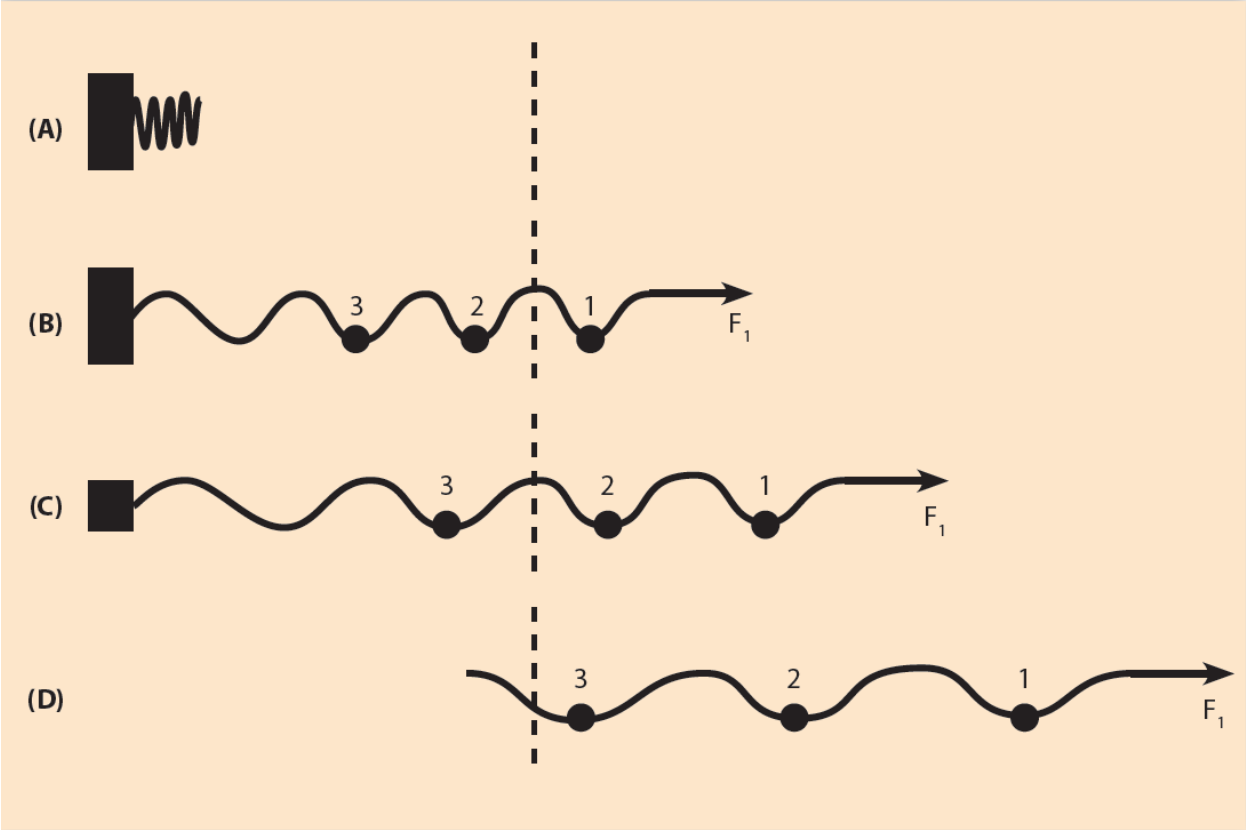
3 The irreversibly expanding spring approximation

With the development of novel technological methods like superresolution imaging of stochastic optical reconstruction microscopy (STORM) it is now possible to measure the geometric modifications of Hox clusters during Hox gene expression [12-14]. It was thus found that the Hox clusters during gene activation are gradually elongated across the 3' to 5' axis of the cluster. This observation supports the hypothesis that the Hox clusters behave like an expanding elastic spring. BM predicts this behavior which is further supported by experimental evidence [15-17] (Fig.2). Following an increase of the morphogen gradient (from head to tail) the pulling force **F** increases and the number of extruded Hox genes increases accordingly.

A) The traditional approach

The mechanical properties of the expanding spring follow Hooke's physical law which states that, for a wide range of pulling forces, the elongation of the spring is proportional to the measure of the force **F** which is applied to one of the spring's ends (Fig.3). At the other end of the spring, the spring is fastened. For the spring's proper function, besides the pulling force **F**, the spring's fastening is equally important. In the case of the mouse HoxD cluster, the **gene regulatory region** of the cluster plays the role of the spring's fastening domain. For simplicity, the role of the 'stiffness' of the spring is ignored together with the local interactions of the constituent chromosomal configuration [18]. In the normal case of animals with *wildtype* development, the spring is completely fastened. If no force is applied on the spring, the spring remains uncharged.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2



[Figure 3]

Elastic spring expansion under a small pulling force (schematic). A) In the uncharged spring no force is applied. The spring is compacted next to the fastening region. B) A small force **F1** is applied to the right end of the spring. The spring fastening is complete and the spring expands slightly. Small sphere 1 moves to the right beyond the dashed line where gene activation occurs. C) The spring fastening is reduced (smaller rectangle) and the spring slides further to the right (two small spheres pass to the activation region. D) The fastening is completely reduced and under the same force **F1** all 3 small spheres 1, 2, 3 move to the activation region.

The traditional tools and methods to explore the genetics of gene clusters are the chemical analyses of the biomolecules involved. This traditional methodology in Hox gene research is combined with genetic engineering techniques of DNA excision or duplication and the subsequent biomolecular analysis of the expression modifications of the neighboring Hox genes (see e.g. [19, 20]).

According to experiments performed with the above traditional techniques, it is now established that the regulatory elements of the mouse Hox clusters are posteriorily located upstream of the cluster even beyond gene Evx2 (Fig.1B). A detailed search has explored this

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

upstream area. Partial excisions of this area cause modifications of the hox gene expressions compared to the *wild type* expressions [21].

If the force **F** is weak, the spring will be slightly shifted while the spring fastening is complete (Fig.3B). If the fastening is partly relaxed, by removing part of the gene regulatory region, the same weak force **F** will slide the spring further (Fig.3C). If the fastening is completely removed, **F** will shift the spring even further (Fig,3D). The above picture leads to a BM prediction: the total removal of the gene regulatory region causes automatically the shifting of the whole cluster in the gene activation region while no gradual Hox gene expressions should be observed.

The above expectation is a BM prediction in retrospect: amazingly, as early as 1999 T. Kondo and D. Duboule observed this phenomenon since Hoxd4 and Hoxd10 expressions appear earlier 'at a time corresponding to that of Hoxd1 appearance **as if temporal collinearity disappeared'** [21, p.414]. Therefore, the expanding spring approximation of BM directly relates the complete deletion of the gene regulatory elements of the HoxD cluster with the disappearance of TC as a result of the early movement of the whole cluster inside the transcription factory domain.

B) A novel approach involving conserved non-coding elements (CNE)

The last twenty years, a novel powerful weapon has been added to the tools exploring the properties and activation of Hox gene clusters. Besides the protein coding genes in the genome, there are more than 30.000 RNA elements in the human genome which do not code any proteins (ncRNA). More than 1000 of these non-coding RNAs are persistently conserved from generation to generation [22]. Lately, the conserved DNA non-coding elements (CNE) and their long non-coding RNAs (lncRNA) have been intensively studied by many groups with the help of sophisticated numerical analysis methods [23,24]. Their findings are impressive. The numerous CNEs are preserved for more than 400 million years of evolution. The size of these CNEs varies, it can reach more than 300 bp. The CNEs play different roles in normal development and disease, coordinating spatial-temporal gene expression in both embryos and grown up animals. A CNE that has been thoroughly studied is *Hotair* which is a lncRNA located between HOXC11 and HOXC12 in the vertebrate chromosome 12 [24]. Its location in the posterior domain of the HOXC cluster is compatible with the hypothesis that CNEs can be involved in the creation of the pulling forces of BM. In what follows, it is indicated that CNEs play an important role

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

particularly in the expanding spring approximation. It is further assumed that several CNEs are located in the fastening posterior domain of the vertebrate Hox clusters incorporating gene Evx2.

In the early studies of CNEs it was reported the striking differences in structure and function of the CNEs occuring in various vertebrates, in particular in Humans and mice [25]. For instance the human *Hotair* represses *in trans* HOXD gene expressions while deletion of mouse Hotair *in vivo* does not affect the Hoxd transcriptions. Later studies did not confirm that the deletion of mouse Hotair has a detectable effect on Hoxd gene expressions *in vivo* [26]. The situation is not clear since the mouse Hotair knock out causes derepression of HoxD genes [27]. Even if Hotair shows low sequence conservation in several vertebrates, it has been noticed that many ncRNAs are conserved in structure although not conserved in sequence [28].

In view of the above findings, an International Bioinformatics Group in collaboration with the Imperial College team headed by B. Lenhard have recently performed a comprehensive analysis of the regulatory roles of Hotair *in cis* and *in trans* on Hox clusters [24]. Nepal et al propose that at the second round of whole-genome duplication, HOTAIR expression is correlated positively with HOXC11 *in cis* and negatively correlated with HOXD11 *in trans* [24]. They compared human and zebrafish CNEs and identified a 32-nucleotide long CNE conserved across the vertebrates. They characterized this long CNE as the ancestral sequence of the ancestral HoxC/D cluster. Their conclusion is that a lncRNA locus functions at the DNA and RNA levels regulating genes both *in cis* and *in trans* [24] .This is a challenging hypothesis to be further tested. More specifically, this analysis indicates that HOTAIR expression regulates positively human HOXC11 in *cis* and negatively HOXD11 in *trans* [24].The number of the conserved elements varies, depending on the cluster copies during the whole genome duplications. Therefore, a certain pulling force will cause a variable shift (sliding) of the cluster toward its telomeric end and the starting time of Hox gene expression will be accordingly modified.

In the expanding spring approximation (approach A), the regulatory gene domain controls the cluster fastening [15,17]. In the novel approach B, the fastening is achieved with the action of a set of CNEs. Note that both approaches lead to comparable results (see Section 6).

As stated in the Introduction, in the case of *Xenopus laevis* M. Kondo *et al.* analyzed the Hox gene expressions and concluded that the initiation of these expressions deviates from the normal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

paralogue order of Hox gene expressions, therefore they violate TC [5]. Following the WGD the number of CNEs fastening the Hox clusters may vary so that the total effect on Hox cluster expansion differs for the various clusters. According to approach B, the initiation and duration of gene activation is reshuffled. This could explain the violation of TC in the L and S homologs of the *Xenopus leavis* as observed by M. Kondo *et al.* [5].

4 Symmetries in a (finite) linear ordering

Symmetry is a very broad concept with many facets covering physical, mathematical, aesthetic and philosophical aspects. Many appropriate definitions have been proposed [29 - 31]. For the present purpose it is preferred the compact definition of Wilczek in the form of aphorism: symmetry is '**Change without change'** [31]. For example, consider a circle in a plane and a perpendicular axis passing through its center. Any rotation around this axis is an operation that moves all points of the circle to some other points of the circle so that the circle remains invariant. This very simple and obvious example is the start of a far going line of thoughts that Emmy Noether followed in 1918 and discovered a fundamental principle of Mathematics and Physics with Philosophical repercussions. Noether started with Classical Mechanics and proved exactly that a physical system obeying a symmetry law is necessarily followed by a conserved quantity. As an example, Noether proved that a physical law symmetrical under spatial rotations (freedom to choose the orientation of the coordinate system) leads to the **conservation of angular momentum** [30, 31]. A simple but rigorous exposition of this Noether Theory (NT) is found in [32]. (For the significance of Symmetries and NT see the Appentix I).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

[Figure 4]

A self-similar design of a BRANSLEY LEAF: Each branch of the leaf is similar to all other branches differing only in their spatial scale (bigger or smaller).

Here, of particular interest is the symmetry of multiscale objects or phenomena. A class of such objects are the **fractals** [33]. The symmetry of fractals is called self-similarity and characterizes the property of an object being similar to its part. In Nature and Life self-similar entities are quite frequent. For example, self-similar is the pattern of the branching pattern of the blood vessels in the lung or the multiscalar Bransley leaf [33] (Fig.4). In the original (theoretical) formulation of self-similarity the scale variation is continuous (and infinite), whereas in HGC only two spatial scales are involved- the embryonic and the Hox gene cluster scales. Nevertheless, in this primitive case of Symmetry it is tempting, according to NT, to expect a primitive conservation of some corresponding quantity. In the common ancestor Hox gene cluster, by inspection, the complete ordering of Hox genes (Hox1, Hox2, Hox3,…,Hox13) is conserved (Fig.1a). For vertebrates, in the homologue Hox clusters (HoxA, HoxB, HoxC, HoxD) the ordering is not

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

complete: some genes are missing while the Pg order is preserved. It is believed that together with the WGD and the evolutionary process some genes faded out (gene loss) and all four homologue clusters cooperated with specific roles.

C. amphioxus is the closest relative of vertebrates and a direct descendant of the vertebrate ancestral existing before WGD. *C. amphioxus* possesses a 14th hox gene without any gene loss therefore it is suitable to study the evolutionary history of vertebrates[34]. Furthermore, a numerical comparison study of *C. amphioxus* and mouse Hox clusters showed that BM is consistent with vertebrate Hox clusters adopting more involved organization as a tool to develop more complex body structures during evolution [35].

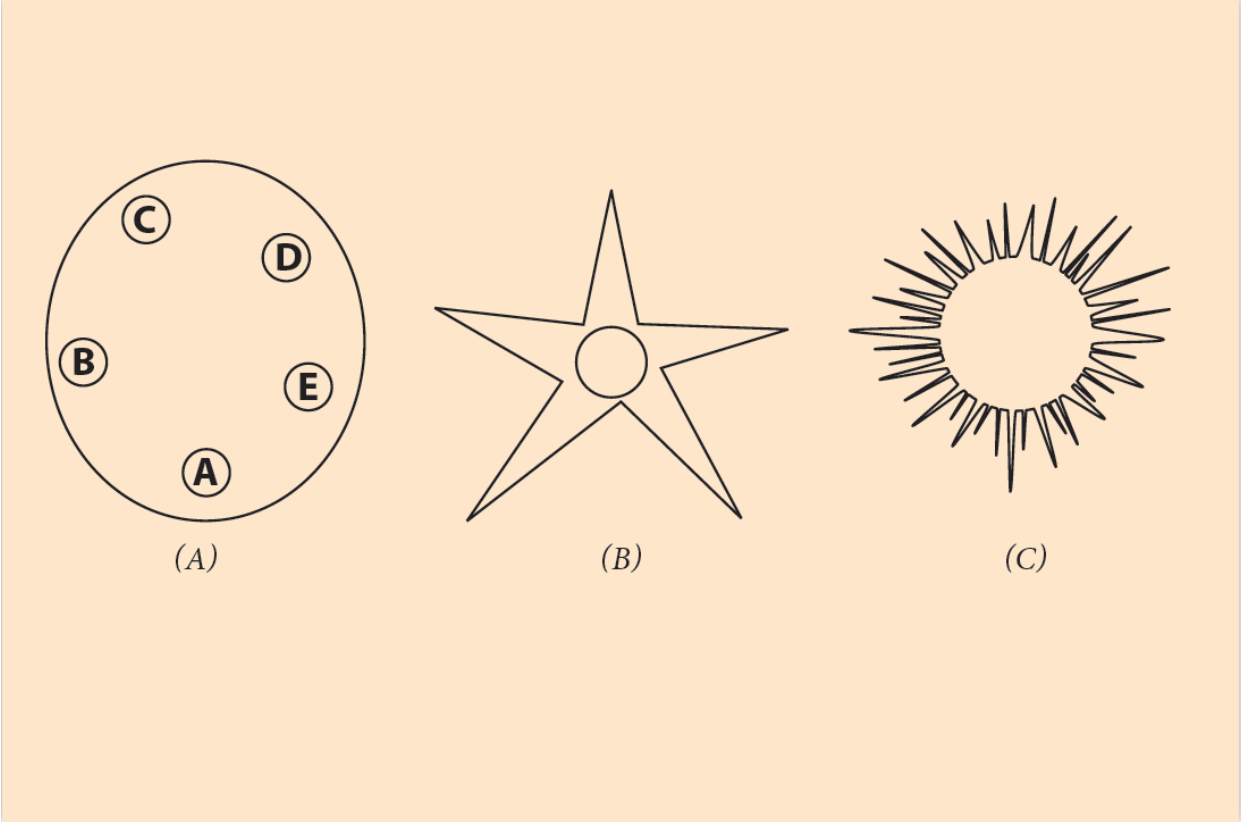
By inspection of the existing *wild type* vertebrate Hox gene data, a partly conserved Hox ordering is observed which follows the rule: the Hox gene order is **irreversibly increasing** like a 'ratchet' even if some genes are missing. For instance, the *wild type* Pg ordering [1, 2, , , 5, 6,...] is allowed. In contrast, the ordering [1, 2, , ,6, 5,…] is forbidden because it represents a gene reversal mutation which is not allowed [36]. The abnormal spontaneous mutant of *Antennapedia* in the *Drosophila* is an example where, in the location of antennas, legs are growing as a result of an abnormal reversal of the corresponding Hox genes [37]. These mutations are named *Homeotic.*

5 Symmetries in a circular gene ordering

The Symmetries of Section 4 underlie the description of organisms whose embryos grow along the A/P axis retaining the same pattern in their adult life e.g. arthropods or vertebrates. The embryos of these 'directly' developing animals look like miniatures of the grown up organisms. Besides these animals there is a large variety of invertebrates which, at the very early stages of embryogenesis, are 'indirectly' developing a larva next to the body of the embryo where the pattern of the grown up animal is formed. This body pattern differs substantially from the initial embryo organization along the linear A/P axis [38-41]. *Holopneustes purpurescens* is a typical indirectly developing sea urchin whose larva organization is circular (Fig.5 a,b).

The first genome sequencing data of the echinoderms were published in 2006 and it was followed by many others [39, 40].The sequencing was unexpected and several models have been proposed to explain the data [40 - 41]. However, many questions still remain unanswered [42].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2



[Figure 5]

Circular organization of echinoderms at different developmental stages

A) Holopneustus purpuratus larva with 5 podia. B) An adult starfish with 5 podia. C) A juvenile sea urchin.

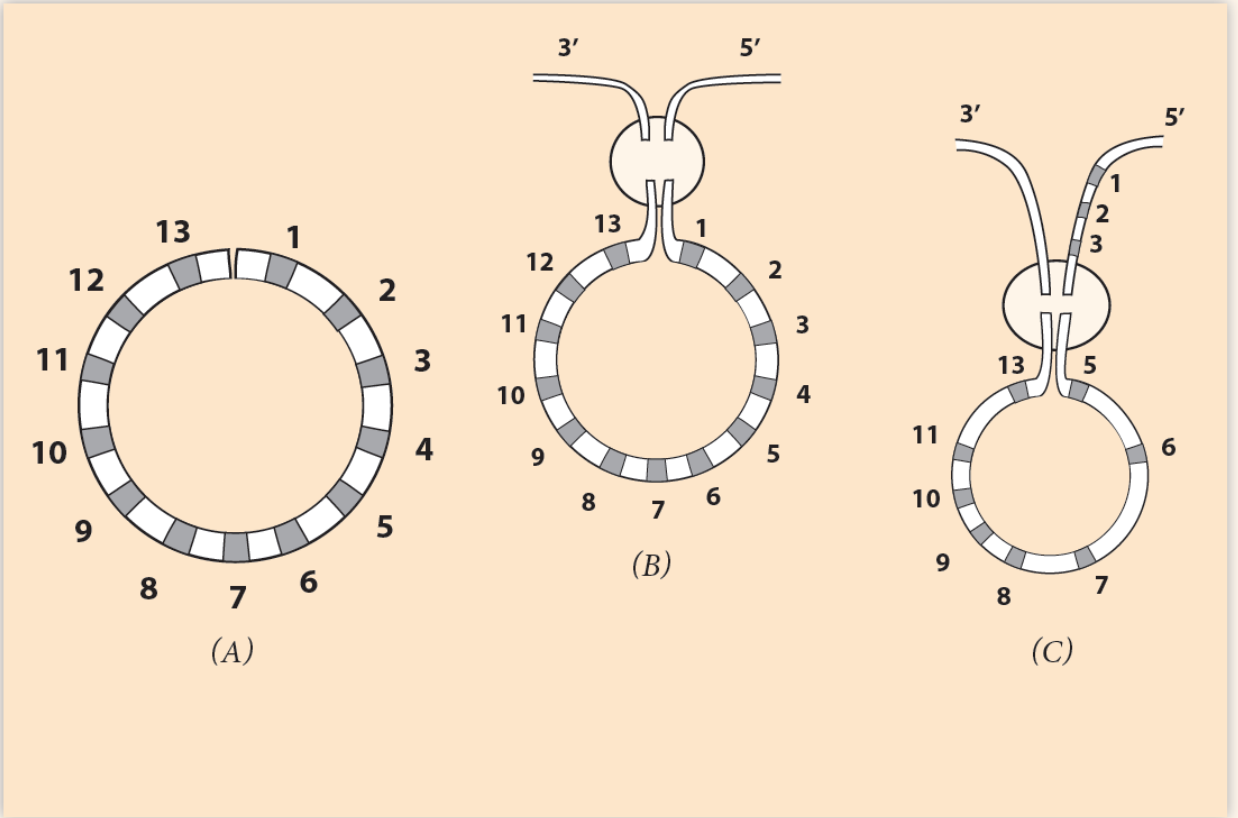
In what follows, another ordering mechanism is proposed based on the Symmetries of the preceding Section. The initial linear embryo organization along the A/P axis is transformed into a circular reorganization at the larva stage. According to NT and the primitive self-similarity of the system, it is expected that the Hox gene cluster is also transformed into an analogue circular organization at the larva stage [42] (Fig.5). Double strand break (DSB) is a mechanism used in experiments of DNA rearrangements to cure serious illnesses like cancer but it is also a tool in spontaneous DNA reshufflings leading to evolutionary novelties. In a recent review both experimental and spontaneous DSB are extensively treated [43]. In hundreds of million years of evolutionary history, DSB played a critical role in the pathway from simple ontological structures to more complex entities.

In Fig.6 a diagram is depicted where, at the larva stage, the Hox genes of a sea urchin cluster

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

are bended forming a loop, conforming to the circular organization of the ontological units of the embryo (Fig.5A). Schematically, the last Hox13 at the posterior end 5' of the cluster is approaching Hox1 at the anterior end 3' in the area where the DSB occurs (in the elliptic disc). In every Hox gene of the circle, a 'circular' identity is imprinted different from the initial linear identity.

If Hox1 is connected to the 3'end of the flanking chromosome (and Hox13 to the 5' end) no novel DNA sequence is created. Although a circularly symmetric invertebrate, *A. planci* retains the Hox ordering of directly developing animals. This is due to the insertion of the cluster in the flanking DNA: Hox1 is connected to the 3' end and Hox13 to the 5'end of DNA. In such insertion no novelty is created (Fig. 6).



[Figure 6]

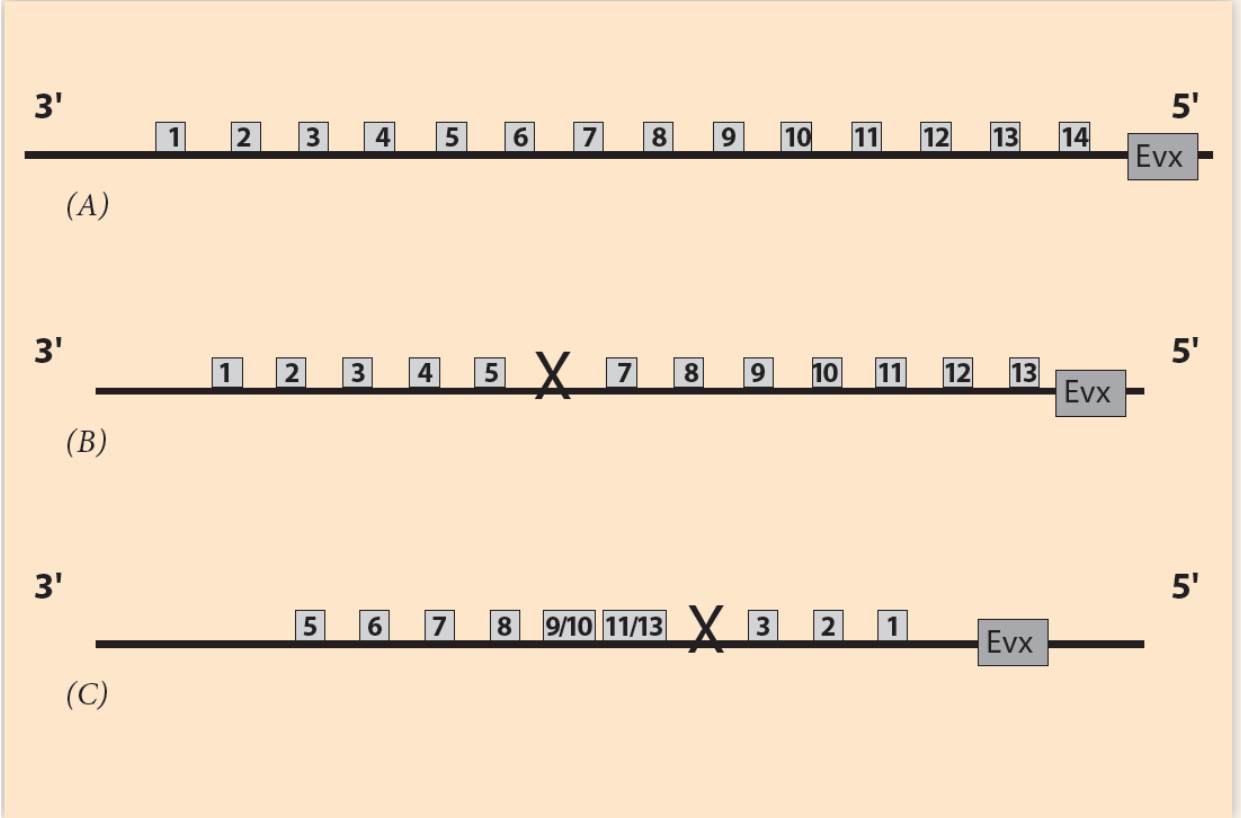
Circular ordering of Hox genes. (Adapted from S. Papageorgiou (2016) Current Genomics).

(A) The linear DNA is bent, and the two ends of the Hox cluster come close together. (B) In the encircled domain, the ends Hox1 and Hox13 of the cluster are connected to the 3' and 5' end of the flanking chromosome. If Hox1 is attached to the 3' end and Hox13 to the 5' end, the produced linear arrangement is the normal one and could represent the observed *A. planci* gene ordering. (C) If Hox5 is

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

connected to the 3' end and Hox13 to Hox3 on the flanking chromosome, the linear ordering is the *sea urchin* Hox cluster arrangement.

In contrast, a novelty will arise if Hox1 is connected to the 5' end of the chromosome (and Hox13 to the 3' end) (Fig.6C). This is the case of the *sea urchin* where the gene order is the result of an abnormal Hox cluster insertion in the flanking DNA (Hox1 to the 5'end and Hox13 to the 3' end) and a novelty is created (Fig.6C). Furthermore, a second DSB can occur at the missing Hox4 location (between Hox3 and Hox5). The two ends of the remnant Hox cluster (Hox5 and Hox13) are correspondingly attached to the two ends 3' and 5'of the flanking DNA (Fig.6C). This leads to the final *sea urchin* gene ordering shown in Fig.7C.



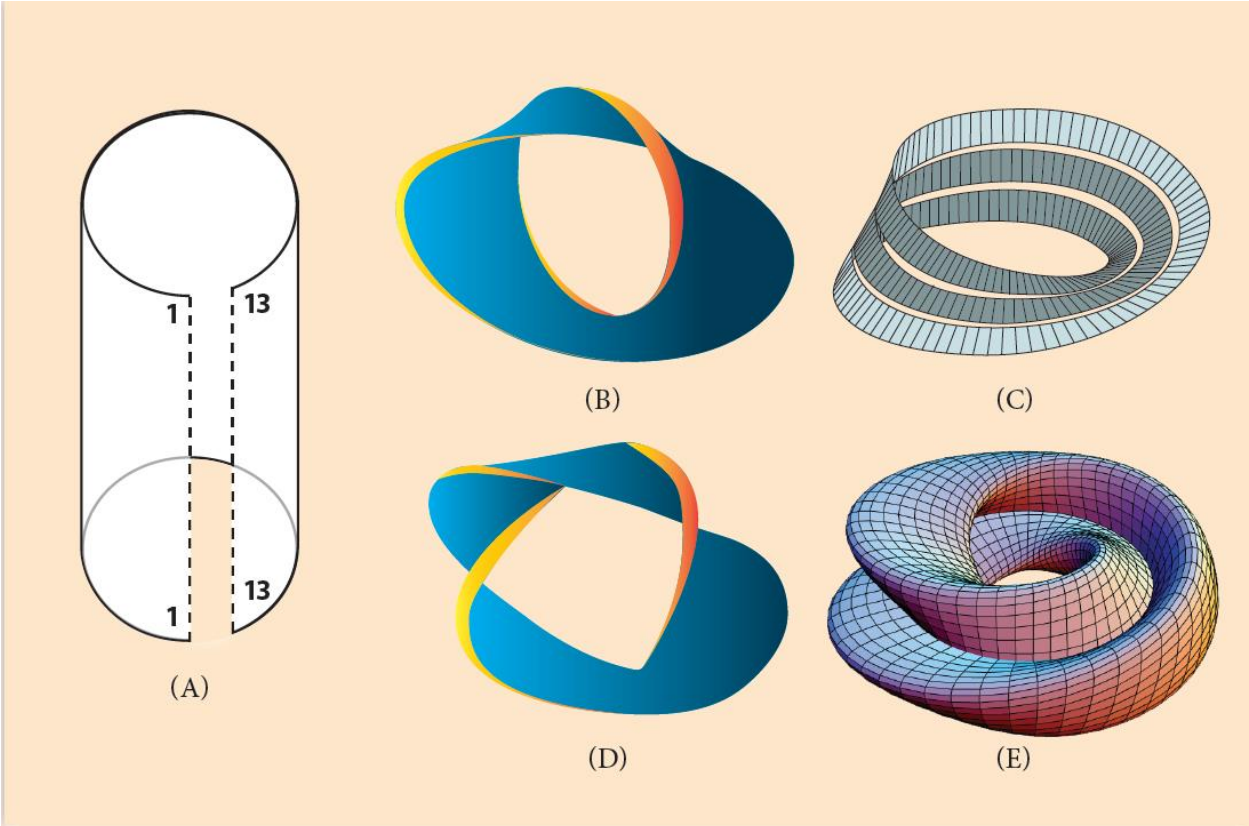
[Figure 7]

Comparison of Hox gene ordering of *Amphioxus*, *A.planci and sea urchin*

A) *Amphioxus* gene cluster has a 14 th Hox gene without any gene loss. B) *A. planci* gene cluster with one gene loss at Pg6 where a DSB occurs. C) *Sea urchin* gene cluster with one gene loss at Pg4 position of DSB.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

At this point it is more realistic to represent the linear Hox cluster ordering of Fig.1 and Fig.6 as a two-dimensional strip. This dimensional change causes a deep topological modification. If this strip is bended so that the two ridges come close to each other, a cylindrical surface will be formed (Fig.8A) [44]. The bending around is performed for 360^0 and the Hox genes obtain a new circular identity. If the cylindical surface is opened the sustem returns to its original two-dimensional strip while the genes retain their circular identity. At the same time another operation can be performed. If one of the strip's edge is twisted 180^0 and then joined to the other edge of the strip, a two-dimensional loop will be created embedded in the 3-dimensional space. The created loop is an endless one-sided surface called **'Moebius strip'** (Fig.8B) [44]. The 2-dimensional twisting of the strip is an extension of the one-dimensional (abnormal) connection of Hox1 to the 5'end of the flanking DNA (and Hox13 to the 3'end) as shown above in Fig. 6.



[Figure 8]

Moebius strip constructions (Adapted from Wikipedia). A) A two-dimensional strip is turned around with the two edges 1/1 and 13/13 coming close to each other. If edge 1/1 is connected to the 3' end of the flanking DNA and edge 13/13 to the 5' end of DNA (look at the one-dimensional connection of (Fig. 5B)). The gene order returns to the normal ordering of Fig. 6B (*A. planci*) [45]. B) A two-dimensional strip is turned around 360^0 and one 180^0 twisting (a Moebius circle). C) a two-dimensional strip turned

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

around 360^0 3 times and one twisting. D) A two-dimensional strip twisted several times. E) a Moebius torus.

This procedure (of bending and twisting) may be extended to 3-dimensional surfaces. A sequential turning around 360^0 with one twisting creates a multiple Moebius strip in the shape of a ring the so-called 'Moebius torus' (Fig.8C). If additional twists are performed more complex structures are created suitable to host several invertebrate genomes with variable symmetries (Fig. 6) [44]. As an example, a Moebius torus can accommodate a sequence of Hox clusters in the chromosome or the genome of a starfish with five podia (Fig.5B).

6 Conclusion

The second half of 20^{th} century was an exciting time for Developmental Biology. Although the discovery of DNA structure was a historical turning point for all branches of Biology, the understanding of many local molecular mechanisms was not equally satisfactory. For instance in 1969, the pivotal work of L.Wolpert and the 'French flag Problem' were quite vague and abstract with no molecular verification[46, 47]. Nevertheless, the notion of morphogens and their gradients were fruitful in guiding research to the right direction. Since then, spectacular technological advancements hand in hand with basic research have promoted our knowledge to an unbelievable depth. For instance, the contemporary status of Development and more specifically the level of our knowledge of Hox genes and the role they play in both Biology and Medicine.

In Section 3 two distinct approaches (A, B) are presented for the Hox gene expressions.The formulations, although looking different, they produce admissible results (not contradicting to experimental evidence). Eventually, the two approaches are equivalent. It is important to explore this possibility by trying (if possible) to formulate an experimental setup where the two approaches predict diverging results. The experimental confirmation of one of them would refute the other(s).

It has been suggested that TC is the 'principal constraining force' keeping Hox clusters in a compact organization [48]. In the case of a complete deletion of the regulatory region, BM

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

predicts that the Hox cluster becomes completely loose, therefore it can move like a freely moving body [49]. A small pulling force (e.g. at a very early stage of activation) initiates the automatic cluster sliding toward the transcription factory domain. As a consequence, no gradual activation of the hox cluster is possible and TC disappears [21] (see Section 3). The above BM prediction leads to a daring hypothesis for the absence of temporal collinearity in *Drosophila*. The mechanism of TC disappearance in *Drosophila* could be the same mechanism predicted above by BM in combination with the evolutionary destruction pathway from an ancestral Hox cluster of 'organized type (O)' [4] to a Hox cluster lacking TC [21].

In Fig. 1 the distances between Hox genes and their expressions are depicted as small squares and circles. This is only schematic and only the 3'end of the gene expressions are clearly observed. The 5' end of these expressions is smeared out so that, in many cases there is an overlap of the cluster gene expressions. It has been further noticed that in cells where such an overlap occurs, the expression intensity of a Hox gene follows its Pg order. For instance, in the same cell the expression intensity of Hox11 gene is stronger compared to the intensity of Hox10. This phenomenon is called 'quantitative collinearity'[9]. (For the overall significant role of Pgs see the Appendix II).

An application of BM could be useful in Cancer research. As stressed above, in the normally developing vertebrate embryos, the pulling forces elongate gradually the Hox cluster [17,49]. It has been observed in many cases that, overexpression of Hox clusters are concurrent with myelodisplastic syndrome [50]. In other cases, overexpression of HoxA and HoxD were found in ovarian cancer with unknown Aetiology [51]. Here a start of an explanation of this phenomenon is proposed: partial or total deletions of the regulatory region of Hox clusters can cause abnormal cluster elongations [17]. These abnormal elongations are related to Hox gene overexpressions. With the plethora of data accumulated in the existing **Big Data Bases** it is probably a matter of 'clever digging' in these data to confirm (or reject) the correlation between mutations in the regulatory region of Hox clusters on one hand, and on the other the abnormal Hox cluster elongations combined with overexpressions and specific forms of malignancy [51].

References

1. Lewis E.B. A gene complex controlling segmentation in *Drosophila*. Nature. 1978;276:565–570. doi: 10.1038/276565a0. [PubMed] [CrossRef] [Google Scholar]

2. Sasaki H., Hogan B.L. Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. Development. 1993;118:47–59. [PubMed] [Google Scholar]

3. Izpisua-Belmonte J.C., Duboule D. Murine genes related to the Drosophila AbdB homeotic genes are sequentially expressed. EMBO J. 1991:10:2279-2289.

4. Duboule D. The rise and fall of Hox gene clusters. Development. 2007;134:2549–2560. doi: 10.1242/dev.001065. [PubMed] [CrossRef] [Google Scholar]

5. Kondo M., Yamamoto T., Takahashi S., Taira M., Comprehensive analyses of Hox gene expression in *Xenopus laevis* embryos and adult tissues. Develop. Growth Differ. 2017:59:526-539.

6. Zhu K., Spaink H.P., Durston A.J. Collinear Hox-Hox interactions are involved in patterning the vertebrate anteroposterior (A-P) axis. PLoS ONE. 2017;11:e0175287. doi: 10.1371/journal.pone.0175287. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

7. Durston A.J. Vertebrate Hox temporal collinearity: does it exist and what is its function? Cell Cycle 2019:18:523-530.

8. Papageorgiou S. A physical force may expose Hox genes to express in a morphogenetic density gradient. Bull. Math. Biol. 2001;63:185–200. doi: 10.1006/bulm.2000.0211. [PubMed] [CrossRef] [Google Scholar]

9. Papageorgiou S. Pulling forces acting on Hox gene clusters cause expression collinearity. Int. J. Dev. Biol. 2006;50:301–308. doi: 10.1387/ijdb.052034sp. [PubMed] [CrossRef] [Google Scholar]

10. Papageorgiou S. Physical forces may cause Hox gene collinearity in the primary and secondary axes of the developing vertebrates. Dev. Growth Differ. 2011;53:1–8. doi: 10.1111/j.1440-169X.2010.01218.x. [PubMed] [CrossRef] [Google Scholar]

11. Papageorgiou S. A biophysical mechanism may control the collinearity of Hoxd genes during the early phase of limb development. Hum. Genomics. 2009;3:275–280. doi: 10.1186/1479-7364-3-3-275. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

12. Noordermeer D., Leleu M., Schorderet P., Joye E., Schabaud F., Duboule D. Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. eLIFE 2014, 3.

13. Fabre P.J., Benke A., Joye E., Huynh T.H.N., Manley S., Duboule D. Nanoscale spatial organization of the HoxD gene cluster in distinct transcriptional states. Proc. Nat. Acad. Sci. 2015: 112:13964-13969.

14. Fabre P.J., Benke A., Manley S., Duboule D. Visualizing HoxD gene cluster at the nanoscale level. Cold Spring Harb. Symp. Quant. Biol. 2015;80:9–16. doi: 10.1101/sqb.2015.80.027177. [PubMed] [CrossRef] [Google Scholar]

15. Papageorgiou S. Comparison of models for the collinearity of Hox genes in the developmental axes of vertebrates. Curr.Genomics 2012: 13:245-251.

16. Alexander T., Nolte C, Krumlauf R. Hox genes and segmentation of the hindbrain and axial skeleton. Annu Rev Cell Dev Biol 2009. doi: 10.1146/annurev.cellbio.042308.113423.

17. Papageorgiou S. Abnormal elongations of HOX Gene clusters may cause cancer. Frontiers in Cell and Developmental Biology 2018: 6:25.

18. Mallo M., Alonso C.R. The regulation of Hox gene expression during animal development. Development 2013:140:3951-3963.

19. Tarchini B., Duboule D. Control of Hoxd genes' collinearity during early limb development. Dev. Cell. 2006;10:93–103. doi: 10.1016/j.devcel.2005.11.014. [PubMed] [CrossRef] [Google Scholar]

20. Parker H.J., Bronner M.E., Krumlauf R. The vertebrate Hox gene network for hindbrain segmentation. Bioessays. 2016;38:526–538. doi: 10.1002/bies.201600010. [PubMed] [CrossRef] [Google Scholar]

21. Kondo T., Duboule D. Breaking collinearity in the mouse HoxD complex. Cell. 1999;97:407–417. doi: 10.1016/S0092-8674(00)80749-7. [PubMed] [CrossRef] [Google Scholar]

22. Washieti S., Hofacker I.L., Lukasser M., Huttenhofer A., Stadler P.F. Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. Nature Biotechnology. 2005: 23:1383-1390.

23. Polychronopoulos D., King J.W.D., Nash A.J., Tan G., Lenhard B. Conserved non-coding elements: developmental gene regulation meets genome organization. Nucleic Acids Research. 2017: 45:22:12611-12624.

24. Nepal C., Taranta A., Hadzhiv Y., Pundhir S., Mydel P., Lenhard B., Muller F., Andersen J.B. Ancestrally duplicated conserved noncoding element suggests regulatory roles of HOTAIR in *cis* and *trans*. iScience. 2020: 23(4):101008.

25. Schorderet P., Duboule D. Structural and functional differences in the long non-coding RNA *Hotair*in mouse and human. PLoS Genetics, 2011: 7:5 e1002071.

26. Amandio A.R., Necsulea A., Joye E., Mascrez B., Duboule D. *Hotair* is dispensible for mouse development. PLOS Genetics. 2016. 12:12.

27. He S., Liu S., Zhu H. The sequence, structure and evolutionary features of HOTAIR in mammals. BMC Evol. Biol. 2011. 11:102.

28. Li L., Liu B., Wapinski O.L. *et al.* Targeted disruption of *Hotair* leads to homeotic transformation and gene –repression. *Cell Rep.* 2013. 5(1):3-12.

29. Weyl H. Symmetry. Princeton University Press, (1952), Princeton, USA.

30. Feynman R.P, Six not-so- easy pieces. Basic Books, 1989, New York.

31. Wilczek, F. A Beautiful Question. Penguin Books, (2015) New York, USA.

32. Marinho, R.M. (2006) Noether's theorem in Classical Mechanics revisited. arXiv:physics/0608264v1

33. Mandelbrot, B.B. (1982) The Fractal Geometry of Nature. W.H. Freeman & Company. New York, USA.

34. Minguillon C. *et al.*(2005), No more than 14: the end of the amphioxus Hox cluster. Int. J. Biol. Sci. 1(1) 19-23.

35. Almirantis, Y.: Provata, A.: Papageorgiou, S. (2013), Evolutionary constraints favor a Biophysical Model explaining Hox gene collinearity. Current Genomics 14, 279-288.

36. Papageorgiou, S. (2020) Hox gene collinearity may be related to Noether Theory on Symmetry and its linked conserved quantity. J- Multidisciplinary Scientific Journal, 3(2), 151-161.

37. Gehring, W.J. Master control genes in Development and Evolutiom. **(**1999) Yale University Press, New Haven, USA.

38. Arenas-Mena, C.; Martinez, P.; Cameron, R. A.; Davidson, E. H. (1998) Expression of the Hox gene complex in the indirect development of a sea urchin. Proc. Nat. Acad. Sci. 95, 13062-13067.

39. Mooi, R.; David, B. (2008) Radial symmetry, the anterior/posterior axis and echinoderm Hox genes.Annu. Rev. Ecol. Evol. 39, 43-62.

40. Cameron, R.A. *et al.* (2006) Unusual gene order and organization of the sea urchin Hox cluster. J. Exp. Zool. B. Mol. Dev. Evol. 306, 45-58.

41. Byrne, M.; Martinez, P.; Morris. V. (2016) Evolution of a pentameral body plan was not linked to translocation of anterior Hox genes: the echinoderm HOX cluster revisited. Evolution & Development 18.2, 137-143.

42. Papageorgiou, S. (2016) Hox Gene Collinearity: From A-P patterning to radially symmetric animals.Curr. Genomics 17, 444-449.

43. Hanscom, T,: McVey M. (2020) Regulation of Error-Prone DNA Double -Strand-Break repair and its impact on Genome Evolution. Cells 9(7): 1657.

44. Steenrod, N. (1999) The topology of Fibre Bundles. Princeton University Press, Princeton, USA

45. Baughman K.W. *et al.* Genomic organization of Hox and ParaHox clusters in the echinoderm, Acanthaster planci. Genesis (2014) doi: 10.1002/dvg.22840.

46. Wolpert, L. (1969) Positional information and the spatial pattern of cellular differentiation. J. theor. Biol. 25: 1-47.

47. Gordon N.K. : Chen, Z.: Gordon, R.:Zou, Y. (2020). French flag gradients and Turing reaction-diffusion versus differentiation waves as models of morphogenesis. doi: 10.1016/j.biosystems.2020.104169.

48. Garstang, M,:Ferrier, D. (2013). Time is of the essence for ParaHox homeobox gene clustering. BMC Biol 26: 11:72.

49. Papageorgiou, S. (2017) Physical Forces may cause the HoxD gene cluster elongation. Biology 6(3): 32.

50. Xu, F. *et al* (2016). Genomic loss of EZH2 leads to epigenetic modifications and overexpression of the HOX gene clusters in myelodysplastic syndrome.Oncotarget doi: 10.18632/oncotarget.6992.

51 Kelly, Z *et al* (2016). The prognostic significance of specific HOX gene expression patterns in ovarian cancer. Int J Cancer 139:1608-17. doi: 10.1002/ijc.3020.

List of Abbreviations

BM Biophysical model
CT Chromatin territory
CNE Conserved Non-coding Elements
DSB Double Strand Break
HGC Hox Gene Collinearity

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

ICD Interchromosome Domain
NT Noether Theory
SC Spatial Collinearity
TC Temporal Collinearity
TF Transcription Factory
WGD Whole Genome Duplication

APPENDIX I

A new Paradigm: HOX Clusters like Atomic Nuclei

The concept of Atoms introduced by Demokritos is a (theoretical) abstract construction referring to the extremely small pieces of mass that cannot be further divided into smaller constituents. At the turning of the 19th to 20th century scientists thought that the Demokritos concept of atoms were the objects they could not further divide so they named them 'atoms' as we even now call the atom of Hydrogen, the Carbon atom etc. In the following hundred years, huge smashing accelerating machines were invented which could split the atoms and their nuclei into their constituent subunits which were given exotic names like quarks, charmed particles, bosons etc. The big accelerating machines are very few in the world because they are extremely expensive. A European instrument called Large Hadron Collider (LHC) is installed at CERN in Geneva. The quarks, charmed, colored particles or other elementary particles (EP) are the contemporary 'atoms' of Demokritos. Who knows what the EPs will be hundred years from now.

One basic theoretical method in the study of the above EP is the exploration of their Symmetries and Noether's theory is indispensable. A multidimensional Supersymmetry is believed to apply to EP at the realm of quanta and the associated conserved quantity is the 'heavy boson' responsible for the creation of the mass of EP (J. Iliopoulos 'Aux origins de la masse: particules elemantaires et symmetries fondamentales' 2014, Editions EDP Sciences, Paris, France). After many laborious efforts, this 'heavy boson' was discovered at the CERN Laboratory in 2012. At the macroscale embryonic scale, a primitive self -similarity symmetry is associated with a remnant NT where the Hox genes of a cluster are irreversibly ordered.

APPENDIX II

Correlation of **Pythagorian theory and Paralogy group**

The paralogy group (Pg) of a Hox cluster is the set of integer numbers corresponding to the ordered set of Hox genes of the cluster. E.g. the common ancestal group of Hox genes defines the paralogy group [Pg1, Pg2, Pg3, , Pg12 , Pg13] of vertebrates. Every Pg has its own

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 March 2021 doi:10.20944/preprints202103.0450.v2

characteristic features that can be traced in several other phyla. In the process of Evolution a Hox gene can change up to extinction (gene loss) but it cannot be substituted in its place by any other Hox gene. The vertebrate homologue clusters (HoxA, HoxB, Hoxc, HoxD) cooperate in the evolutionary pathway for the development of novelties (more involved structures). For the 4 homologue vertebrate clusters, the number of Pg is 39 and this number combines with the13 missing Hox genes (gene losses). It has been noticed that DSB occurs at the location of gene losses and subsequently the gene cluster is inserted in the flanking DNA. These properties are reminiscent of the analysis of the integer numbers by Pythagoras. Number Theory is a high standing branch of Mathematics for all times. The Pythagorian theorem is the cornerstone of Number Theory. In a generalized version this theorem takes the form of an equation:

$$\mathbf{a^n + b^n = c^n} \qquad \text{(A)}$$

where a, b, are the smaller sides of an orthogonal triangle and c its hypotenuse. All numbers a, b, c and n are integers. For n=1 the solution is trivially simple: a + b= c.

For n = 2, Eq.(A) is the proper Pythagorian theorem and only few sets of integers (a, b, c) can satisfy this equation. The triplet (a = 3, b =4, c =5) was discovered by Pythagoras himself as a solution of Eq.A. It is the basic solution while any other solution is a multiple triplet (e.g. a=2x3, b= 2x4, c=2x5). It was the first historical Paradigm of a rigorous mathematical 'proof '. Astonishingly, for any other integer value of n (bigger than n=2), no solution of Eq. (A) exists with integer values of the triplet (a,b,c). Pythagoras extended the study of integers to the frequencies of vibrating strings and established the correlations of these frequencies with the length of the strings and the resulting acousting 'harmony' for some ratios of string lengths in integer numbers. His credo was: **All Things Are Number** [31].

For the vertebrate four homologue Hox clusters, the total number of Pgs is 39 and the gene losses 13. It is conjectured that this set 39+13 = 52 obeys a correlation (code) *à la* Pythagoras to explain both the pattern of Hox cluster activation and the position of the breaking point.