



Noether's Theorem as a Metaphor for Chargaff's 2nd Parity Rule in Genomics

Yannis Almirantis¹ · Astero Provata² · Wentian Li³

Received: 26 January 2022 / Accepted: 18 May 2022 / Published online: 15 June 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

In the present note, the genomic compositional rule largely known as ‘Chargaff's 2nd parity rule’ (asserting equimolarity between Adenine–Thymine and Guanine–Cytosine in any of the two DNA strands) is regarded in association with Noether's theorem linking symmetries with conservation laws in physics. In the case of the genome, the strict physical and mathematical prerequisites of Noether's theorem do not hold. However, we conclude that a metaphor can be established with Noether's theorem, as inter-strand symmetry concerning DNA functionality engenders specific features in genome composition. Inversely, when inter-strand symmetry does not hold, the corresponding quantitative relations fail to appear. This association is also considered from the point of view of the existence of emergent laws and properties in evolutionary genomics.

Keywords Chargaff's 2nd parity rule · Noether's theorem · DNA inter-strand symmetry · Conservation laws · Emergent laws in evolutionary biology

In 1915, mathematician Emmy Noether formulated and proved a theorem (the subsequently termed Noether's theorem) which states that every differential (thus continuous) symmetry that holds in a physical system has a corresponding conservation law (Noether 1918). For example, in systems which are symmetrical under translation in space, momentum is conserved and in systems endowed with rotational symmetry, angular momentum is conserved. Also, symmetry in time (that is, when any experiment produces the same results independently of the moment that it starts) according to this theorem leads to energy conservation.

Several other symmetries and corresponding conserved quantities according to this theorem have been explored as modern physics kept developing, but these cases are outside the scope of the present description. The interested reader can find systematic outlines of Noether's theorem, which stress historical (Byers 1998), more technical (Marinho 2006) and other aspects as well (Brown and Holland 2004; Wigner 1954). Moreover, Kosmann-Schwarzbach (2011) has written an excellent book on Noether's theorem and its physical applications.

Several years after the formulation of Noether's theorem, the discovery of the structure of DNA and of its role as a carrier of genetic information gave rise to the new fields of molecular biology and genomics. It was already well known before the determination of the double helical structure of DNA and has been fundamental for its discovery, that when total DNA is analyzed, i.e., if both strands are considered, exact equimolarity holds between complementary nucleotides, i.e., between adenine (A) and thymine (T) or guanine (G) and cytosine (C): $A = T$ and $G = C$, where A, T, G, C denote the molar fractions of the four bases. This is the so-called *interstrand basepairing rule* (BPR).

Early in the history of molecular biology, Erwin Chargaff, one of its pioneers, and his collaborators (Rudner et al. 1968), discovered that the above relations also hold approximately for the composition of long stretches of

Handling editor: **Anthony Poole**.

✉ Yannis Almirantis
yalmir@bio.demokritos.gr

¹ Theoretical Biology and Computational Genomics Laboratory, Institute of Bioscience and Applications, National Center for Scientific Research “Demokritos”, 15341 Athens, Greece

² Statistical Mechanics and Dynamical Systems Laboratory, Institute of Nanoscience and Nanotechnology, National Center for Scientific Research, “Demokritos”, 15341 Athens, Greece

³ The Robert S. Boas Center for Genomics and Human Genetics, The Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, USA

single-stranded genomic DNA. While BPR is clearly related to and explained by the complementary structure of double-stranded DNA, the observation of Chargaff et al. posed a considerable challenge to the developing field of genomics. We will briefly review here the cause behind these quantitative relationships as analyzed in a seminal work by Sueoka (1995, 1996). At the same time, answer to this problem was also provided by Lobry (1995a, b) independently, using a slightly different approach. The condition met when mutation and selection are equally effective on both strands (i.e., when strand-equivalence holds) is termed *first parity rule* (PR1). In the works of Sueoka and Lobry cited above is shown that when PR1 holds true and the substitutional dynamics has sufficient time to reach its equilibrium state, then, equalities $A = T$ and $G = C$ become valid for the composition of any of the two single DNA strands. This is termed Chargaff's *second parity rule* (PR2). As stated, prerequisite for PR2 and thus for the intrastrand $A = T$ and $G = C$ relations to hold is the absence of biases in mutation and selection between the two strands.

In terms of an interplay between symmetry and the corresponding conserved quantities we may consider the case where the double-stranded DNA is symmetric under the following substitutional modalities. Let us call $P^I(A \rightarrow T)$ the base substitution rate that reflects the combined effect of a mutation rate (m) and a selection coefficient (s) from A to T in strand I and accordingly for all other possible base substitution events in strands I and II (see Sueoka 1995). Namely, in the case we describe here the following base-substitution equalities hold: $P^I(A \rightarrow T) = P^I(T \rightarrow A)$, $P^I(C \rightarrow G) = P^I(G \rightarrow C)$, $P^I(A \rightarrow G) = P^I(T \rightarrow C)$, $P^I(A \rightarrow C) = P^I(T \rightarrow G)$, ... and similarly for the whole set of 12 substitution rates. See (Lobry 1995a) and "Appendix". This symmetry, as expressed by the above equalities, allows to reduce the number of free base-substitution constants from 12 to 6, see (Sueoka 1995, 1996). This is equivalent to PR1 and constitutes between-strands *symmetry* in the DNA structure. Subsequently, the role of the *conservation* here is played by the PR2 intrastrand $A = T$ and $G = C$ relationships, and consequently, for purines (P_u) and pyrimidines (P_y), $P_u = P_y = 0.5$ holds.

To further explore the relation between PR1 and PR2, we recall in the "Appendix" the substitution matrix, which takes into account the aforementioned rate relations (Lobry 1995a). The eigenvectors and eigenvalues of this matrix are useful for determining the asymptotic nucleotide frequencies. Due to the properties of the substitution matrix the largest eigenvalue is unity and corresponds to the asymptotic state (Perron-Frobenius theorem). The components of the corresponding eigenvector correspond to the four nucleotide asymptotic concentrations (molecular fractions). If the PR1 conditions are taken into account, then the resulting asymptotic frequencies lead to $A = T$, $G = C$

and consequently $P_u = P_y = 0.5$, i.e., to PR2. Therefore, the symmetry conditions PR1 lead to equal frequencies of purines and pyrimidines. The PR2 equalities are known to be conserved in most genomes (except for animal mitochondria) as earlier stated and is going to be discussed in the sequel. This symmetry—conservation interplay could be seen as an analogue of Noether's theorem in genomics.

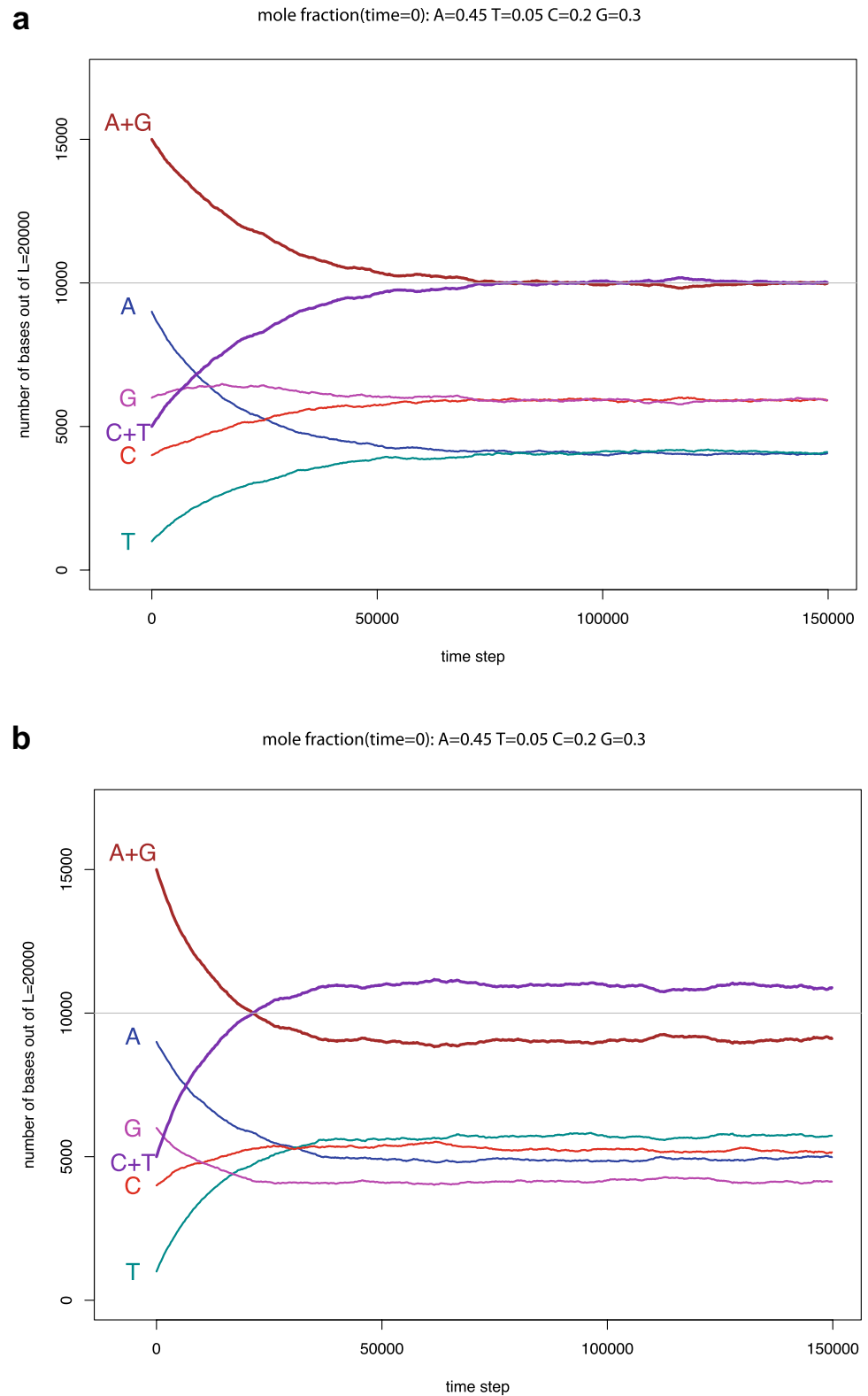
As an exemplary simulation, we consider in Fig. 1 the case of evolution of an initial random sequence of length $L = 2 \times 10^4$ nucleotides (nts), where the initial values of molecular fractions are: $A = 0.45$, $T = 0.05$, $C = 0.2$ and $G = 0.3$. In Fig. 1a no biases in mutation and selection between the two strands of DNA are assumed, i.e., PR1 holds. The inverse condition is met in Fig. 1b. In accordance with the theory exposed above, in Fig. 1a the asymptotic emergence of the compositional properties denoted as PR2, i.e. $A = T$, $C = G$ and $P_u = P_y = 0.5$, are clearly visible. This is not the case when PR1 is not respected, see Fig. 1b. Similar results are obtained for different initial molecular fractions (not shown here).

As we may infer from the above analysis, deviations from PR2 might be used for the detection of differences in the functionality of the two DNA strands, differences between their roles in genetic signaling or between the local environments where they are exposed. Cumulative presentation of the excess quantities $(A - T)/(A + T)$ and $(G - C)/(G + C)$, the so-called relative nucleotide skews, has been shown to generate V- or Λ -shaped patterns along the linear chromosome coordinate, which is suitable for following such trends in DNA composition (Grigoriev 1999).

While PR2 is obeyed by most genomes of organisms or organelles when their entire length is considered, notable exceptions also exist. These exceptions concern parts of genomes, particularly halves of circular bacterial chromosomes delimited by the 'origin of replication' (*ori*) and its diametrical point 'terminus of replication' (*ter*). Frank and Lobry (2000) developed *Oriloc*, a successful online computational tool for the positioning of *ori* and *ter* in bacteria, based on the above observation. While, in a first approximation, deviations from equivalence between leading and lagging strand during DNA replication are at the origin of departures from PR2, transcription also contributes to it, as the strands of opened DNA are being exposed to different environmental influences during the transcription as well as during the replication process (Beletskii and Bhagwat 1998; Seplyarskiy and Sunyaev 2021). In that way asymmetric mutational pressures between leading and lagging strand or between coding and complementary strand can be shaping the nucleotide composition violating PR1 (and thus PR2) through both processes (Francino and Ochman 2001; Fijalkowska et al. 1998; Nikolaou and Almirantis 2005; Necşulea and Lobry 2007).

Two additional points might be mentioned here in brief:

Fig. 1 a Here, in a computer simulation, the A (blue), T (green), G (magenta), C (red), Pu [A + G] (brown) and Py [C + T] (purple) molecular fractions (initial values are arbitrary chosen) for a DNA molecule of 2×10^4 nts, are monitored in evolutionary time. Interstrand parity in the mutation rates is supposed (i.e., PR1 holds). One observes that (i) approximate equalities $A \cong T$ & $G \cong C$ tend to be established and (ii) $Pu \cong Py \cong 0.5$ are reached asymptotically. This is an illustration of the emergence of the PR2 ‘conservation’ relationships as a result of the interstrand ‘symmetry’ relations dictated by PR1. **b** In a numerical set-up analogous to the previous ones (color convention as above) when mutation rates do not obey PR1, then PR2 relations fail to establish. Thus, the resulting DNA molecule disposes a heavy (purine-rich) and a light (pyrimidine-rich) strand, similarly to the genome of animal mitochondria (Color figure online)



- First, in bacteria, not only mononucleotide skews (i.e., deviations from PR2) but also skewed first-neighbor preferences are shown to follow V- or Λ -shaped patterns, as a result of deviations from PR1 related to muta-

tional dynamics, as we have systematically investigated in a previous work (Apostolou-Karampelis et al. 2016). Moreover, as analytically discussed therein, PR1 symmetry produces *intra-strand first-neighbor preferences*

of the PR2 type, under conditions of inter-strand equality of context-dependent substitution rates (see Fig. 1 in the aforementioned work).

- Second, in the eukaryotic chromosome, a more complicated pattern emerges when single-nucleotide skews are considered, the so-called ‘factory-roof motif’, due to the coexistence of numerous origins of replication. This is again a consequence of functional asymmetry between the two strands (Touchon et al. 2005).

While entire bacterial and eukaryotic chromosomes mostly follow PR2, in some cases, organellar genomes deviate from what dictates PR2 for their entire length (Mitchell and Bridge 2006). In a systematic review of these small and peculiar genomes (Nikolaou and Almirantis 2006), it is shown that animal mitochondria deviate considerably and systematically from PR2. Indeed, this compositional feature (here, of whole mitochondrial genomes) is shown to relate to asymmetry in structure and function between the two strands: mitochondrial DNA strands differ in their content in purines and pyrimidines, thus named heavy and light strands. Moreover, in vertebrate mitochondria, genome duplication occurs through the *rolling circle* mode, meaning that both strands are replicated continuously starting from two different single strand origins. This replication dynamics leads to two strands being exposed to different mutational dynamics in all their length, which represents a complete deviation from PR1, and consequently leading to a whole-length violation of PR2 in these organellar genomes.

We have to keep in mind that symmetry considerations between DNA strands involved in the explanation of PR2 are not of the type described in the prerequisites of Noether’s theorem (Noether 1918; Byers 1998; Wigner 1954; Kosmann-Schwarzbach 2011). They consist of, as briefly stated here and analyzed in (Sueoka 1995, 1996; Lobry 1995a, b), absence of biases in mutation and selection between the two strands, thus they are mainly of a functional origin. This makes any parallelism between Noether’s theorem and the emergence of PR2 a metaphor rather than a direct one. However, in both: (a) the case of symmetry causing conservation in Noether’s theorem and (b) RP1 causing RP2 discussed here, the cause—effect link can be mathematically derived exactly. The *prerequisite*—some kind of symmetry or equivalence between strands—and its *consequence*—the compositional constraints $A = T$ and $G = C$, which hold true in the measure of validity of the strand equivalence—brings the connection between PR1 and PR2 close to the essence of Noether’s theorem. Although, $PR1 \rightarrow PR2$ relationship lies outside the range of the mathematical prerequisites of Noether’s theorem, we may consider PR2 as the consequence of PR1, just like *symmetry acts as a cause* and *conservation as a consequence* in Noether’s theorem. Additionally, the PR1 relations have, as a consequence that $P_u = P_y = 0.5$, which, technically, sounds close to a typical ‘conservation’ property. Therefore, the symmetry

expressed by PR1 leads to several forms of quantitative relationships: compositional variables of specific information-bearing macromolecules are constrained to remain almost equal, and the two strands are led to always conserve equal percentages of purines and pyrimidines.

Although applications of Noether’s theorem come principally from the field of Physics, here, symmetry engenders quantitative relationships within the realm of biology. Another aspect that makes peculiar this analogy is that here relative nucleotide skews only approach zero in a time infinity limit (i.e., conservation-like relationships are approximate), while deviations from equality (skews) are studied as bearing valuable information for complex biological procedures which may involve several mutational but also selectional biases (Grigoriev 1999; Seplyarskiy and Sunyaev 2021; Francino and Ochman 2001; Fijalkowska et al. 1998; Nikolaou and Almirantis 2005; Necşulea and Lobry 2007). In entirely different contexts, in developmental biology and brain dynamics, Noether’s theorem relation to biology have been reported recently (Papageorgiou 2020; Bilteanu et al. 2017).

Overall, the observation of PR2 in genomic sequences is a case where symmetry in function between DNA strands, i.e. (a) equivalence in the mutational pressure they undergo and (b) equivalence in their biological roles (thus almost-equality of constraints due to evolutionary selection), has compositional consequences. The interplay between (interstrand) *symmetry* and (intrastrand) ‘*conservation*’ relationships has parallels with Noether’s theorem, although, as stated, it cannot be considered to be a direct application of this theorem. Instead, we suggested that this seminal theorem may be seen as *a metaphor* for PR2 and this brings out a conceptual analogy between physics and biology. This metaphor discussed in the present note may be viewed in the perspective described by Eugene Koonin in his article: “Are There Laws of Genome Evolution?” (Koonin 2011) and in other related studies as well (West et al. 1997; Guiot et al. 2003; Li 2011; Marchal 2015; Militaru and Munteanu 2013; Longo and Montévil 2011). In Koonin’s work (Koonin 2011), a selection of the most conspicuous such universal regularities is presented, which share the status of ‘laws of evolutionary genomics’, in the same sense that ‘law’ is understood in modern physics. They include: log-normal distribution of the evolutionary rates between orthologous genes; power-law-like distributions of membership in paralogous gene families and node degree in biological “scale-free” networks; negative correlation between a gene’s sequence evolution rate and expression level (or protein abundance); distinct scaling of functional classes of genes with genome size. For the original references on these relationships and detailed discussion see (Koonin 2011).

We would like to stress Koonin’s remark that the universal regularities discussed in his article do not appear to be shaped by selection but rather are emergent properties of gene ensembles. Similar, in that respect, is the phenomenon examined herein: according to the analysis of Sueoka (1995) and Lobry (1995a, b), PR2 results as a consequence of the

mutational dynamics and of the whole cellular activity. This is the case, provided that selection or functional divergence between strands either act weakly or their impact is strand-independent, when the entire genome is considered (e.g., in typical bacterial genomes). Therefore, the cause-effect link between PR1–PR2 approaches the status of a physical law—and thus can be seen as an analogue of Noether’s theorem—as far as dependence on natural selection is relaxed. This means that PR2 is non-adaptive in the sense that emergent properties reviewed by Kooning—e.g., scale-free networks of biochemical reactions etc.—are also non-shaped by selective pressure. All these regularities can be considered as constituting an ensemble of ‘physical laws’ met in biological systems.

In the last two decades, much work has been done investigating the genomic composition looking for PR2-type relations between oligo-nucleotides (k -mers) and their reverse complement with very interesting results. In several works, alternative mechanisms additional to the PR1 \rightarrow PR2 (i.e., additional to the compositional intra-strand relations generated by inter-strand symmetry) have been formulated. It is out of the scope of the present work to systematically review this research field. Indicatively, we recall below some characteristic cases and their connections to the PR1–PR2 relationship.

In the work of Afreixo and co-workers, see e.g. (Afreixo et al. 2013, 2016), the appearance of PR2 relations has been explored in the case of k -mers for k up to 10. This research group has examined the statistical significance of the measured compositional traits, focusing on the ‘exceptional symmetry’, meaning symmetry beyond that expected under an independent nucleotide assumption (i.e., under no neighbor preference). They conclude that up to $k=7$, characteristic deviations occur from what is expected on the grounds of randomness. They also observed a complex compartmentalization of the genome involving: k -mers abundances, patterns of occurrence of PR2 and diversification of genomic functionality including protein-coding and non-coding regions.

Forsdyke and co-workers have developed an explanation for the establishment of PR2 relationships for both single-nucleotides and k -mers, proposing that: “Chargaff’s second parity rule reflects the evolution of genome-wide stem-loop potential ...”, see (Bell and Forsdyke 1999) and references given therein. This approach brings attention to a factor which might contribute to the finally measured PR2 relations, provided that a given sequence does exhibit considerable stem-loop potential. However, this is expected mainly in the protein-coding regions, thus leaving unexplained the PR2 equalities that hold in the noncoding, especially within the long eukaryotic chromosomes. Later on, Forsdyke, following the same trail of ideas, has also elaborated on the alignment-free phylogenetic classification (Forsdyke 2019), which represents a

particularly fertile line of thought, although far from the subject of the present research note.

Albrecht-Buehler (2006) suggested that: “... inversions and inverted transpositions could be a major contributing if not dominant factor in the almost universal validity ...” of PR2 for single- and oligo-nucleotides. Although they describe a mechanism certainly contributing to PR2, the degree of its contribution to the final picture remains an open question.

Both approaches of ‘stem-loop potential’ and of ‘inversions and inverted transpositions’ cannot explain, even in principle, either the V- and Λ -shaped skew patterns of single nucleotides (Grigoriev 1999), or the corresponding similar patterns found in the first-neighbor preferences (Apostolou-Karampelis et al. 2016) in bacteria. Also, they cannot account for the ‘factory-roof motif’ in eukaryotic chromosomes (Touchon et al. 2005).

We have to stress here that inter-strand symmetry in function (PR1) can account not only for single-nucleotide PR2, but also for equalities in the frequency of occurrence of k -mers. This last property happens on the basis of strand equivalence in contextual (neighbor-dependent) substitution rates, which then engenders PR2-type relations for k -mers. In a previous work (Apostolou-Karampelis et al. 2016), one of us and co-workers have analyzed the link between inter-strand symmetry and PR2 relations for k -mers. The appearance of accurate PR2 equalities for dimers does depend on equivalence between strands, which concerns known molecular mechanisms: (a) No or minimal strand-biases in the function of PolIII α -subunit during RNA transcription is necessary for the appearance of PR2 relations for k -mers. Also, (b) the contextuality (dependence on neighboring nucleotides) in transcription-coupled repair, and (c) in other repair-mechanisms or genomic dynamics in general is examined and shown to influence the values of frequency of occurrence of specific dinucleotides. Therefore, in (Apostolou-Karampelis et al. 2016) it was shown that *PR2 for k -mers is ultimately reduced to inter-strand symmetry considerations*.

Rosandić and co-workers (Rosandić and Paar 2014; Rosandić et al. 2016, 2019) relate the appearance of PR2 to the inter-strand mirror symmetry in 20 symbolic purine—pyrimidine symmetry quadruplets of trinucleotides (direct, reverse complement, complement, and reverse) mapped to double-stranded genome. Their description is compatible with the above dynamical approaches and emphasizes the link between the compositional relations (PR2) at all k -mers’ lengths with intrer-strand symmetry (PR1).

Closing this research note on the connection between Chargaff’s PR2 relations and Noether’s theorem, we should stress that the context of this type of work is very wide, see e.g. (Byers 1998; Wigner 1954; Kosmann-Schwarzbach 2011; Wigner et al. 1969; Monod 1978), where views from

physics are progressively incorporated into main stream biology, especially in the study of the interplay between the genome and its function.

Appendix: The Properties of the Substitution Matrix

Let us call T the “matrix of the substitution rates”, which has the following form:

$$T = \begin{pmatrix} P_{AA} & P_{AT} & P_{AC} & P_{AG} \\ P_{TA} & P_{TT} & P_{TC} & P_{TG} \\ P_{CA} & P_{CT} & P_{CC} & P_{CG} \\ P_{GA} & P_{GT} & P_{GC} & P_{GG} \end{pmatrix} \quad (1)$$

where $P_{AT} = P(A \rightarrow T)$ represents the substitution from nucleotide A to T and similarly for the other nucleotides. Note that $P_{AA} = 1 - P_{AT} - P_{AC} - P_{AG}$. The element P_{AA} represents the case that no substitution of A toward the other nucleotides takes place. Similar relations hold for the substitutions of nucleotides T , C and G . Equation (1) represents a generic substitution matrix, without any particular assumptions. The matrix T acts on the vector V^i , which is composed by the initial, “i”, molecular fractions of nucleotides:

$$\vec{V}^i = \begin{pmatrix} V_A^i \\ V_T^i \\ V_C^i \\ V_G^i \end{pmatrix} \quad (2)$$

where V_A^i , V_T^i , V_C^i and V_G^i denote the initial molecular fractions of the four nucleotides A , T , C and G . If the matrix T acts on the vector V^i many times, at the asymptotic limit, the final, long time nucleotide concentrations, V^f , is reached.

Let us now impose the aforementioned no-strand biases condition (Chargaff 1st parity rule, PR1), which corresponds to the following relationships, restricting the twelve transition rates to only six (Lobry 1995a):

$$\begin{aligned} P_{AT} &= P_{TA} = x \\ P_{CG} &= P_{GC} = y \\ P_{AG} &= P_{TC} = z \\ P_{AC} &= P_{TG} = u \\ P_{GA} &= P_{CT} = v \\ P_{GT} &= P_{CA} = w \end{aligned} \quad (3)$$

Let us consider first the trace of the substitution matrix T , which is the sum of the diagonal elements. The trace of T is also equal to the sum of the matrix eigenvalues:

$$\begin{aligned} \text{Trace}(T) &= 4 - P_{AG} - P_{AC} - P_{AT} - P_{TA} - P_{TC} - P_{TG} \\ &\quad - P_{CA} - P_{CT} - P_{CG} - P_{GA} - P_{GT} - P_{GC} \end{aligned} \quad (4)$$

This trace remains *invariant* under the substitutions $P_{AT} \rightarrow P_{TA}$, $P_{TA} \rightarrow P_{AT}$, $P_{AG} \rightarrow P_{TC}$, etc., without even the

need of the equalities (3). The trace T is a first invariant quantity, exhibited by the dynamics under the substitution conditions PR1. The pair PR1–PR2 can be regarded in analogy with the pair symmetry—invariance described in the Noether’s theorem.

Regarding the determinant of the matrix T , we note that in the general case it is not invariant under the above mere substitutions. Nevertheless, in the case when PR1 holds, the equalities (3) are valid and the determinant gets also invariant under the above-mentioned substitutions. Note that the determinant is equal to the product of the eigenvalues.

Under conditions PR1, Eq. (3), the substitution matrix T , Eq. (1), reduces to T_s :

$$T_s = \begin{pmatrix} 1 - x - u - z & x & u & z \\ x & 1 - x - u - z & z & u \\ w & v & 1 - y - u - w & y \\ v & w & y & 1 - y - v - w \end{pmatrix} \quad (5)$$

Due to its construction, the stochastic matrix T_s , has a largest eigenvalue $\lambda_{\max} = 1$. This is because each row of T_s sums up to 1. The eigenvector of the transpose matrix which corresponds to eigenvalue λ_{\max} dictates the asymptotic behavior represented by the vector V^f of the asymptotic (final) concentrations (Lobry 1995a, b).

The analysis of the eigenvectors of matrix T_s , and in particular of the components of the eigenvector corresponding to the largest eigenvalue, together with the concentration normalization condition $V_A + V_G + V_C + V_T = 1$, which holds for all times (initial, final and intermediate), lead to the PR2 equalities for the final nucleotide concentrations:

$$\begin{aligned} V_A^f &= V_T^f, \quad V_C^f = V_G^f \text{ and} \\ \text{Pu} &= \text{Py} = 0.5 \end{aligned} \quad (6)$$

The sum of V_A^f and V_G^f constitutes the final concentration of purines Pu , while the sum of V_C^f and V_T^f corresponds to the final concentration of pyrimidines, Py . As stated in the main text, relation (6) represents the almost-equality of Pu and Py experimentally observed in genomes where 2nd parity rule holds overall. In the opposite case, where Eq. (6) is not obeyed the emergence of a light and a heavy chain is observed (e.g., animal mitochondria) as also discussed in the main text.

In relation again to a Noether’s theorem (symmetry-conservation pairs), while the largest eigenvalue ($\lambda_{\max} = 1$) is also invariant in the general case of substitutions, the corresponding eigenvector components V_A^f , V_G^f , V_C^f and V_T^f are not invariant per se. Only when substitutions combined with equalities (3), equivalent to PR1, hold, then the Pu and Py concentrations are also constant and equal within the same strand [Eq. (6), PR2].

Overall, the mere substitutions $P_{AT} \rightarrow P_{TA}$, $P_{TA} \rightarrow P_{AT}$, $P_{AG} \rightarrow P_{TC}$, etc., dictated by the Chargraff's 1st parity rule conserves (a) the trace of the dynamics and (b) the largest eigenvalue $\lambda_{\max} = 1$. In addition, the substitutions combined with the equalities of PR1 lead to equal purine—pyrimidine participation in each strand: $Pu = Py = 0.5$.

Acknowledgements The authors would like to thank Drs. J. Alevizos and A. Mistriotis for helpful discussions.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. The authors have no competing interests to declare that are relevant to the content of this article.

References

- Afreixo V et al (2013) The breakdown of the word symmetry in the human genome. *J Theor Biol* 335:153–159. <https://doi.org/10.1016/j.jtbi.2013.06.032>
- Afreixo V et al (2016) The exceptional genomic word symmetry along DNA sequences. *BMC Bioinform* 17(1):1–10
- Albrecht-Buehler G (2006) Asymptotically increasing compliance of genomes with Chargaff's second parity rules through inversions and inverted transpositions. *Proc Natl Acad Sci USA* 103(47):17828–17833
- Apostolou-Karampelis K, Nikolaou C, Almirantis Y (2016) A novel skew analysis reveals substitution asymmetries linked to genetic code GC-biases and PolIII a-subunit isoforms. *DNA Res* 23(4):353–363. <https://doi.org/10.1093/dnares/dsw021>
- Beletskii A, Bhagwat AS (1998) Correlation between transcription and C to T mutations in the non-transcribed DNA strand. *Biol Chem* 379(4–5):549–551
- Bell SJ, Forsdyke DA (1999) Accounting units in DNA. *J Theor Biol* 197(1):51–61
- Bilteanu L, Casanova MF, Opris I (2017) Symmetry & noether theorem for brain microcircuits. In: Opris I, Casanova M (eds) *The physics of the mind and brain disorders*. Springer series in cognitive and neural systems, vol 11. Springer, Cham. https://doi.org/10.1007/978-3-319-29674-6_6
- Brown HR, Holland P (2004) Simple applications of Noether's first theorem in quantum mechanics and electromagnetism. *Am J Phys* 72(1):34–39. <https://doi.org/10.1119/1.1613272>
- Byers N (1998) E Noether's discovery of the deep connection between symmetries and conservation laws. arXiv [arXiv:physics/9807044v2](https://arxiv.org/abs/physics/9807044v2)
- Fijalkowska IJ, Jonczyk P, Tkaczyk MM, Bialoskorska M, Schaaper RM (1998) Unequal fidelity of leading strand and lagging strand DNA replication on the Escherichia coli chromosome. *Proc Natl Acad Sci USA* 95(17):10020–10025
- Forsdyke DA (2019) Success of alignment-free oligonucleotide (k-mer) analysis confirms relative importance of genomes not genes in speciation and phylogeny. *Biol J Linn Soc* 128(2):239–250
- Francino MP, Ochman H (2001) Deamination as the basis of strand-asymmetric evolution in transcribed Escherichia coli sequences. *Mol Biol Evol* 18(6):1147–1150. <https://doi.org/10.1093/oxfordjournals.molbev.a003888>
- Frank AC, Lobry JR (2000) OriLoc: prediction of replication boundaries in unannotated bacterial chromosomes. *Bioinformatics* 16(6):560–561
- Grigoriev A (1999) Strand-specific compositional asymmetries in double-stranded DNA viruses. *Virus Res* 60(1):1–19
- Guiot C, Degiorgis PG, Delsanto PP, Gabriele P, Deisboeck TS (2003) Does tumor growth follow a “universal law”? *J Theor Biol* 225(2):147–151. [https://doi.org/10.1016/S0022-5193\(03\)00221-2](https://doi.org/10.1016/S0022-5193(03)00221-2)
- Koonin EV (2011) Are there laws of genome evolution? *PLoS Comput Biol* 7(8):e1002173. <https://doi.org/10.1371/journal.pcbi.1002173>
- Kosmann-Schwarzbach Y (2011) The Noether theorems invariance and conservation laws in the twentieth century. SpringerLink (trans: Schwarzbach BE)
- Li W (2011) On parameters of the human genome. *J Theor Biol* 288:92–104. <https://doi.org/10.1016/j.jtbi.2011.07.021>
- Lobry JR (1995a) Properties of a general model of DNA evolution under no-strand bias conditions. *J Mol Evol* 40(3):326–330
- Lobry JR (1995b) Erratum. Properties of a general model of DNA evolution under no-strand bias conditions. *J Mol Evol* 40(3):326–330. *J Mol Evol* 41:680
- Longo G, Montévil M (2011) From physics to biology by extending criticality and symmetry breakings. *Prog Biophys Mol Biol* 106(2):340–347
- Marchal B (2015) The universal numbers from biology to physics. *Prog Biophys Mol Biol* 119(3):368–381. <https://doi.org/10.1016/j.pbiomolbio.2015.06.013>
- Marinho RM (2006) Noether's theorem in classical mechanics revisited. arXiv [arXiv:physics/0608264v13](https://arxiv.org/abs/physics/0608264v13)
- Militaru R, Munteanu F (2013) Symmetries and conservation laws for biodynamical systems. *Int J Math Models Methods Appl Sci* 7(12):965–972
- Mitchell D, Bridge R (2006) A test of Chargaff's second rule. *Biochem Biophys Res Commun* 340(1):90–94
- Monod J (1978) On symmetry and function in biological systems. In: Ullmann A (ed) *Selected papers in molecular biology by Jacques Monod*. Academic Press, Cambridge, pp 701–713
- Necşulea A, Lobry JR (2007) A new method for assessing the effect of replication on DNA base composition asymmetry. *Mol Biol Evol* 24(10):2169–2179
- Nikolaou C, Almirantis Y (2005) A study on the correlation of nucleotide skews and the positioning of the origin of replication: different modes of replication in bacterial species. *Nucleic Acids Res* 33(21):6816–6822
- Nikolaou C, Almirantis Y (2006) Deviations from Chargaff's second parity rule in organellar DNA insights into the evolution of organellar genomes. *Gene* 381:34–41
- Noether E (1918) Invariante Variationsprobleme, *Nachr d König Gesellsch d Wiss zu Göttingen, Math-phys Klasse* 235–257. English trans: M A Travel, *Transp Theory Stat Phys* 1(3):183–207 (1971)
- Papageorgiou S (2020) Hox gene collinearity may be related to Noether theory on symmetry and its linked conserved quantity. *J (MDPI)* 3(2):151–161
- Rosandić M, Paar V (2014) Codon sextets with leading role of serine create “ideal” symmetry classification scheme of the genetic code. *Gene* 543(1):45–52
- Rosandić M, Vlahović I, Glunčić M, Paar V (2016) Trinucleotide's quadruplet symmetries and natural symmetry law of DNA creation ensuing Chargaff's second parity rule. *J Biomol Struct Dyn* 34(7):1383–1394. <https://doi.org/10.1080/07391102.2015.1080628>
- Rosandić M, Vlahović I, Paar V (2019) Novel look at DNA and life—symmetry as evolutionary forcing. *J Theor Biol* 483:109985
- Rudner R, Karkas JD, Chargaff E (1968) Separation of B subtilis DNA into complementary strands 3 direct analysis. *Proc Natl Acad Sci USA* 60(3):921–922

- Septyarskiy VB, Sunyaev S (2021) The origin of human mutation in light of genomic data. *Nat Rev Genet* 22(10):672–686. <https://doi.org/10.1038/s41576-021-00376-2>
- Sueoka N (1995) Intra-strand parity rules of DNA base composition and usage biases of synonymous codons. *J Mol Evol* 40(3):318–325. *J Mol Evol* 42:323
- Sueoka N (1996) Erratum. Intra-strand parity rules of DNA base composition and usage biases of synonymous codons. *J Mol Evol* 40(3):318–325
- Touchon M, Nicolay S, Audit B et al (2005) Replication-associated strand asymmetries in mammalian genomes: toward detection of replication origins. *Proc Natl Acad Sci USA* 102(28):9836–9841
- West GB, Brown JH, Enquist BJ (1997) A general model for the origin of allometric scaling laws in biology. *Science* 276(5309):122–126. <https://doi.org/10.1126/science.276.5309.122>
- Wigner EP (1954) Conservation laws in classical and quantum physics. *Prog Theor Phys* 11(4&5):437–440
- Wigner EP, Seeger RJ, Cohen RS (1969) Physics and the explanation of life. 11 *Boston studies in the philosophy of science*. Springer, Dordrecht, pp 119–132