ABSTRACT: The idea of computational processes, which take place in nature, for example, DNA computation, is discussed in the literature. DNA computation that is going on in the immunoglobulin locus of vertebrates shows how the computations in the biological possibility space could operate during evolution. We suggest that the origin of evolutionarily novel genes and genome evolution constitute the original intrinsic computation of the information about new structures in the space of unrealized biological possibilities. Due to DNA computation, the information about future structures is generated and stored in DNA as genetic information. In evolving ontogenies, search algorithms are necessary, which can search for information about evolutionary innovations and morphological novelties. We believe that such algorithms include stochastic gene expression, gene competition, and compatibility search at different levels of structural organization. We formulate the increase in complexity principle in terms of biological computation and hypothesize the possibility of in silico computing of future functions of evolutionarily novel genes.

KEYWORDS: Biological computation, complexity increase, possibility space

Introduction

In modern literature, the idea of computational processes, which take place in nature, is widely discussed, including the computational notion of life.1 So, the concept of intrinsic computation, which refers to “how dynamical systems store, structure, and transform historical and special information,” is being used.2 The concepts of molecular computation are also actively discussed and implemented.3-5 Molecular computation means the capability of organic molecules to process the information. The authors consider various kinds of molecular operations with DNA sequences as natural computational processes. DNA computing by human is based on DNA structure and on the processes, in which DNA participates in nature.4 Based on these operations, the authors build DNA computers and create new computing paradigms.3,4

In our previous publications6-7 we introduced the notion of the search engine in the space of biological possibilities in relation to the evolutionary origin of morphological novelties. In the present paper, we will study the broader concept of biological computation and compatibility search as the mechanism of the complexity growth in evolution. We will use the terms “biological computation” and “DNA computation” to describe computations done by life, and the terms “DNA computing” and “biocomputing” in relation to computing by human.8

The Space of Biological Possibilities

The substances and structures existing in nature may be considered as realized possibilities.9

Popper6 developed a theory of propensities and the possibility space. According to this theory, the universe is a world of propensities, “an unfolding process of realizing possibilities and of unfolding new possibilities.” The possibility space is growing. All non-zero possibilities will realize themselves in time. Possibilities that have not yet realized themselves have a kind of reality. The future is actively present at every moment, as “a reality in the making.” Causation in a world of propensities is probabilistic causation, and natural laws have probabilistic character.8 Similar ideas continue developing in the scientific literature.7,9

Based on Popper’s theory, we may formulate a concept of the space of biological possibilities, in which evolution occurs. The concept of biological possibility space is connected in some way with “omics” technologies and biological disciplines, which generate and use “big data.” We think that in the course of evolution the search for new structural entities took place in the biological possibility space.7 The evolutionarily new gene, which does not have a function yet, may be considered as the “reality in the making.”

The Increase in Complexity Principle

The capability to increase the degree of aggregation and complexity of the structure is an intrinsic property of the matter. In our earlier publications, we considered the increase of the degree of aggregation and complexity as the fundamental principle7,10,11 although it may realize on short-term/localized scale in cosmological terms. In non-living matter, the increase in structural complexity occurs under the laws of physics (ie, the increase in complexity of the elements in the Mendeleev periodic table). The origin of life is considered a result of chemical evolution.12 Organic molecules originate from mixtures of
simple gases if the energy source is available. In this way, the potential of atoms and simple molecules to generate more complex molecules was realized. Biological evolution is characterized by the further increase in the complexity of biological molecules and evolving organisms. Lamarck formulated the concept of gradation, which is the concept of the gradual evolution of the living matter from simple organization to complex and perfect organization. The increase in complexity of the living matter may give a direction to evolution. Pervasive and long-term evolutionary trends of morphological complexity increase have been described. The increase in complexity is the basis of the concept of progressive evolution (see eg, Shanahan for a review of the idea of evolutionary progress). The increase of complexity may not be adaptive, at least at the initial stages of complexification (reviewed in Kozlov).

Using Popper’s terminology, the evolution of complexity takes place in the possibility space and is determined by pre-existing structures.

**The Anticipatory Role of Evolution at the DNA Level**

Since the origin of the genetic code, the evolution of proteins depended on preliminary evolution at the DNA level, although the latter reckoned with regularities of protein evolution. For example, gene exons generally correspond to protein domains, and exon shuffling corresponds to domain rearrangement.

During progressive evolution, which is characterized by the origin of multigenic morphological novelties, evolution at the DNA level took the lead over the evolution of morphological structures.

So, the coding of new structures may start before the origin of structures themselves. In other words, the information about non-existent structures may be computed at the DNA level.

**DNA Computation of the Information About Future Structures in Progressive Evolution**

Natural computational processes participate in progressive evolution and the increase in complexity of biological structures. DNA computation plays a unique role in the progressive evolution of living organisms because the information about future structures is first computed at the DNA level.

An example of DNA computation that is going on in the immunoglobulin locus of vertebrates shows how the computations in the biological possibility space could operate during evolution.

Due to recombination in immunoglobulin locus (V(D)J-recombination) and subsequent mutations, the multitude of various immunoglobulins is created (10^11 of possible combinations in humans), which corresponds to the multiplicity (the possibility space) of all possible antigens, which vertebrates can meet. The next-generation sequencing (NGS) analysis of antibody diversity has detected 0.5 × 10^9 antibody molecules. It is noteworthy that the authors use the terminology “the observed antibody space” to describe the antibody diversity.

If only one, albeit complex, locus of the vertebrate genome can compute the whole diversity of antibodies and the corresponding diversity of antigens, the evolving vertebrate genome can probably compute the whole space of possibilities for the evolution of vertebrate morphological structures and their adaptations to different environments.

Ohno established the importance of the origin of evolutionarily novel genes for progressive evolution. We suggest that the origin of evolutionarily novel genes and genome evolution constitute the original intrinsic computation of the information about new structures in the space of unrealized biological possibilities. Due to DNA computation, the information about unrealized biological possibilities is generated and stored in DNA as genetic information.

**Gene Competition, Compatibility Search, and Incompatibility Neutralization**

During the individual and evolutionary development of multicellular organisms, the search for entities from the biological possibility space starts from the search in the genetic information space.

In established ontogenies, the retrieval of genetic information is carried out by regulatory mechanisms. Many examples of such mechanisms may be found in the book "The Regulatory Genome." These mechanisms operate in the space of realized biological possibilities.

The peculiarity of the evolutionary search for new entities is that such entities do not exist as yet. In evolving ontogenies, additional search algorithms are necessary, which can search for information about evolutionary innovations and morphological novelties in the space of unrealized possibilities.

The molecular mechanisms, which exist before the origin of regulatory mechanisms, are gene competition and stochastic gene expression. These mechanisms could be used for evolutionary search in the genetic information space of unrealized possibilities.

The basic principle of evolutionary search algorithms is the compatibility of novel and older characters. In connection with genes, we studied this idea in Kozlov. We have shown that in evolving organisms the products of evolutionarily novel genes must be compatible with the products of evolutionarily older genes. Incompatibility between novel genes, which is the consequence of gene competition, is resolved by spatio-temporal disconnection of their products. In multicellular organisms, this is accomplished by cell differentiation and the origin of evolutionarily novel cell types, tissues and organs, and more complex ontogenies. As a result, the divergent functional organization of groups (networks) of compatible genes originated, with incompatible genes acting in different compatibility groups (see Tables 1 and 2 in Kozlov).

Thus, the search for compatible novel genes and gene combinations is an important principle of the evolutionary search algorithm operation in unrealized genetic possibility space.
The evolutionary search algorithm is similar to the binary search algorithm and produces functional genome organization of groups of compatible genes similar to the binary search tree or B-tree.

The evolution of immunoglobulin diversity demonstrates features of the evolutionary innovation search engine. V(D) J-recombination generates the diversity of immunoglobulins by random rearrangement of DNA segments. V(D) J-recombination results in the allelic exclusion of Ig genes, which determines the monospecificity of B lymphocytes and the possibility of clonal selection of antibody-producing cells.26 Similarly, the “one neuron – one receptor” rule is observed in the olfactory system.27 These processes could have evolved due to the search algorithm discussed above.

For the realization of the whole diversity of immunoglobulins encoded by immunoglobulin locus, the other levels or organization (cellular and multicellular) are necessary. The structure of immunoglobulin loci of amphibia and mammals are similar but the diversity of immunoglobulins in mammals is several orders of magnitude higher. This is because amphibia have a small number of cells in their immune organs (reviewed in Kozlov7). So, the evolution of the diversity of immunoglobulins involves interplay between different levels of organization.

**Autonomous Search Engines, Unfolding Possibility Spaces, and Unstable Transitory Forms**

As we discussed earlier,6,10 processes at different levels of organization are relatively independent, and evolution proceeds due to the coincidence of relatively independent events at genomic (DNA), cellular, and multicellular levels. That is, the interplay between different levels of organization has stochastic nature and may generate relatively unstable transitory organisms. Only biologically meaningful coincidences were “frozen” in evolution (“frozen accidents”28) and gave rise to species of organisms with higher complexity and new functions.

Cellular material for the evolutionary search engines in multicellular organisms could be hereditary tumor cells. We discussed this in our previous publications on the possible evolutionary role of heritable tumors (the theory of *carino-evo-devo*7). We have shown that hereditary tumors provided the building material for progressive evolution - additional cell masses, which were functionally not necessary to organisms (autonomous). In tumor cells, evolutionarily novel genes are expressed21,29,30 and compatibility rules are seriously modified (eg, ectopic syntheses8). Stochastic gene expression is characteristic of tumor cells.31-34 That is why hereditary tumor cells, with their epigenetic peculiarities, could play the role of search engines for unrealized genetic information. Hereditary tumors may unfold the possibility space for expression of evolutionarily novel genes and gene combinations, which could result in the origin of new cell types, tissues, and organs in an evolutionary perspective. Populations of tumor-bearing organisms could be transitory forms in the progressive evolution of organisms to higher complexity.6,7,11,35

**Formulating the Increase in Complexity Principle in Terms of Biological Computation**

Thus, the origin of novel organs, the development of which involves constellations of activity of thousands of genes and considerable cell masses, may be explained using the concept of biocomputational processes, as described above. Otherwise, the origin of such complex structures is difficult to imagine.

We suggest a new formulation of an old principle of the increase in complexity in living nature:

The complexity increase in progressive evolution is realized through biological computation of the maximum number of compatible structural entities in evolving lineages of multicellular organisms. Biological computation of complexity increase involves DNA computation in the space of unrealized possibilities; stochastic gene expression and gene competition; compatibility search and incompatibility neutralization at different levels of organization; autonomous search engines and unfolding possibility spaces; unstable transitional forms and “freezing” of biologically meaningful constellations of entities, compatible within the ontogeny of multicellular organisms. The complexity of progressively evolving organisms tends to increase to a maximum31 and can be measured as the number of structural entities from the biological possibility space realized within the ontogeny of the multicellular organism.

The above formulation is a description of progressive evolution and complexity increase as determined by laws of intrinsic causality, or as self-development of biological matter,4 provided that the energy and the material source is available. Darwinian natural selection for adaptation starts when the environment is included in the compatibility search.

The above formulation is a biocomputational description of complexity increase during progressive evolution, which expands the computational notion of life and makes important predictions.

**Predictions of the Biocomputational Concept of Complexity Increase**

After the origin of the genetic code, the evolution of DNA determined the evolution of the living matter. That is why the existing informational network of DNA is not only the storage of information about the previous periods of evolutionary development. It also contains information about the future pathways of evolutionary development, which can be studied by bioinformatic approaches with the help of supercomputers.

We can perform in silico computing of future functions of evolutionarily novel genes (in silico evolution), which are in the stage of their origin (Popper’s reality in the making).

Tumor specifically expressed, evolutionarily novel genes (*TSEEN* genes), predicted by *carino-evo-devo* theory, have been described in the laboratory of the author.21,29,30 In
complexity increase demonstrates features of biological computer science application of the above considerations.

Similar to the idea of quantum computing supremacy, the notion of cellular supremacy in biocomputing is being considered. In article, the authors were looking for supremacy of biocomputing over traditional computers. They were looking for tasks where biocomputing tradition was possible, and for features of biological systems that could be used for supremacy, focusing mostly on the cellular (bacterial) biocomputational processes. Our study extends the biocomputing supremacy considerations to complexity increase in the course of the progressive evolution of multicellular organisms. Progressive evolution resulted in the enormous increase of complexity and functional efficiency of progressively evolving organisms. The biological complexity increase demonstrates the features of biological computation which may be used to reach supremacy: DNA computation of the information about new structures; multi-level structural organization and interplay between different levels of organization; independent evolutionary development at different levels; stochastic mechanisms and "frozen accidents"; compatibility search and neutralization of incompatibilities at different levels of structural organization; autonomous search engines and unfolding possibility spaces; and unstable transitional structures. That is why our findings may be complementary to Grozinger et al and our approach may be useful in studying the supremacy of biocomputing problems.

Conclusion
The origin of novel organs during the progressive evolution of multicellular organisms may be explained using the concept of biocomputational processes in the biological possibility space. The increase in complexity principle formulated in terms of biological computation predicts that information about the future pathways of evolutionary development can be studied by bioinformatic approaches with the help of supercomputers. Biological complexity increase demonstrates features of biological computation that may be used to reach supremacy in biocomputing.

Acknowledgements
The author thanks anonymous reviewers for helpful comments.

Author Contributions
AP Kozlov is an only author.

ORCID iD
AP Kozlov https://orcid.org/0000-0003-4611-1534

Notes
a. The author is grateful to anonymous reviewer for his or her helpful advice on terminology.

b. We may guess if such information already exists in the possibility space, but this question is similar to the question of whether mathematics exists or is invented, and deserves separate consideration.

c. Human genome contains more than 20000 protein-coding genes. Up to 479 morphological characters can be scored that may be an index of vertebrate morphological complexity. Humans have at least 411 differentiated cell types, which may be used as another measure of complexity.

d. The discussion of the present-day status of this problem may be found in our book.

e. Evolutionarily novel genes originate either from pre-existing genes or de novo. A review of the mechanisms of gene origin may be found in our book. The "intrinsic computation" is defined by Feldman et al.

f. The genotype spaces are being discussed in the literature. The concept of morphological and phenotypic space as a set of all possible characters is also used in biological literature.

g. Using the terminology of Niemann, we may say that the products of incompatible genes occupy different possibility spaces within the organism. As shown below, hereditary tumors may provide additional possibility spaces in the evolutionary perspective.

h. Leibniz introduced the idea of compossibility. He thought that the existing world consists of compatible entities. Popp et al, wrote about "exclusive possibilities" and "preferences of the organisms for certain possibilities.

i. Similar to the tendency of the entropy to a maximum in an isolated system.

j. Berg formulated the concept of evolution determined by law, or nomogenesis, which he opposed to the Darwinian theory of natural selection.

k. By self-development we mean the internal capability of living matter to generate more complex structures. A similar notion used in the literature is self-organization.

l. Both studies suggest that the stochastic nature of biological processes can be used in supremacy considerations.

REFERENCES
1. Emmeche C. The computational notion of life. Theoria: Segunda Epoca. 1994;9:30.
2. Feldman DP, McTague CS, Crutchfield JP. The organization of intrinsic computation: complexity-entropy diagrams and the diversity of natural information processing. Chaos. 2008;18:043106.
3. Adleman LM. Molecular computation of solutions to combinatorial problems. Science. 1994;266:1021-1024.
4. Paun G, Rozenberg G, Salomaa A. DNA Computing: New Computing Paradigms. Springer-Verlag; 1998.
5. Grozinger L, Amos M, Gorochowski TE, et al. Pathways to cellular supremacy in biocomputing. Nat Commun. 2019;10:5250.
6. Kozlov AP. Evolution by Tumor Neofunctionalization: The Role of Tumors in the

Matyunina et al, we have shown that orthologs of fish TSEEN genes acquired progressive functions in humans. We plan to use machine learning algorithms to analyze the trajectory of TSEEN gene evolution between fishes and humans. The obtained pattern of gene evolution will be used to predict the future functions of human TSEEN genes. Such computing will be similar to the calculation of unoccupied electron orbitals in physics.

After computing the future functions of human TSEEN genes we will be able to use this information to influence human tumors in which these genes are expressed.

Complexity Increase and Supremacy of Biocomputing
The purpose of this paper was to describe complexity increase in the course of progressive evolution in terms of biological computation. Nevertheless, we cannot miss the possible computer science application of the above considerations.

Similar to the idea of quantum computing supremacy, the notion of cellular supremacy in biocomputing is being considered. In article, the authors were looking for supremacy of biocomputing over traditional computers. They were looking for tasks where biocomputing tradition is possible, and for features of biological systems that could be used for supremacy, focusing mostly on the cellular (bacterial) biocomputational processes. Our study extends the biocomputing supremacy considerations to complexity increase in the course of the progressive evolution of multicellular organisms. Progressive evolution resulted in the enormous increase of complexity and functional efficiency of progressively evolving organisms. The biological complexity increase demonstrates the features of biological computation which may be used to reach supremacy: DNA computation of the information about new structures; multi-level structural organization and interplay between different levels of organization; independent evolutionary development at different levels; stochastic mechanisms and "frozen accidents"; compatibility search and neutralization of incompatibilities at different levels of structural organization; autonomous search engines and unfolding possibility spaces; and unstable transitional structures. That is why our findings may be complementary to Grozinger et al and our approach may be useful in studying the supremacy of biocomputing problems.

Conclusion
The origin of novel organs during the progressive evolution of multicellular organisms may be explained using the concept of biocomputational processes in the biological possibility space. The increase in complexity principle formulated in terms of biological computation predicts that information about the future pathways of evolutionary development can be studied by bioinformatic approaches with the help of supercomputers. Biological complexity increase demonstrates features of biological computation that may be used to reach supremacy in biocomputing.

Acknowledgements
The author thanks anonymous reviewers for helpful comments.

Author Contributions
AP Kozlov is an only author.

ORCID iD
AP Kozlov https://orcid.org/0000-0003-4611-1534

Notes
a. The author is grateful to anonymous reviewer for his or her helpful advice on terminology.

b. We may guess if such information already exists in the possibility space, but this question is similar to the question of whether mathematics exists or is invented, and deserves separate consideration.

c. Human genome contains more than 20000 protein-coding genes. Up to 479 morphological characters can be scored that may be an index of vertebrate morphological complexity. Humans have at least 411 differentiated cell types, which may be used as another measure of complexity.

d. The discussion of the present-day status of this problem may be found in our book.

e. Evolutionarily novel genes originate either from pre-existing genes or de novo. A review of the mechanisms of gene origin may be found in our book. The "intrinsic computation" is defined by Feldman et al.

f. The genotype spaces are being discussed in the literature. The concept of morphological and phenotypic space as a set of all possible characters is also used in biological literature.

g. Using the terminology of Niemann, we may say that the products of incompatible genes occupy different possibility spaces within the organism. As shown below, hereditary tumors may provide additional possibility spaces in the evolutionary perspective.

h. Leibniz introduced the idea of compossibility. He thought that the existing world consists of compatible entities. Popper et al, wrote about "exclusive possibilities" and "preferences of the organisms for certain possibilities.

i. Similar to the tendency of the entropy to a maximum in an isolated system.

j. Berg formulated the concept of evolution determined by law, or nomogenesis, which he opposed to the Darwinian theory of natural selection.

k. By self-development we mean the internal capability of living matter to generate more complex structures. A similar notion used in the literature is self-organization.

l. Both studies suggest that the stochastic nature of biological processes can be used in supremacy considerations.

REFERENCES
1. Emmeche C. The computational notion of life. Theoria: Segunda Epoca. 1994;9:30.
2. Feldman DP, McTague CS, Crutchfield JP. The organization of intrinsic computation: complexity-entropy diagrams and the diversity of natural information processing. Chaos. 2008;18:043106.
3. Adleman LM. Molecular computation of solutions to combinatorial problems. Science. 1994;266:1021-1024.
4. Paun G, Rozenberg G, Salomaa A. DNA Computing: New Computing Paradigms. Springer-Verlag; 1998.
5. Grozinger L, Amos M, Gorochowski TE, et al. Pathways to cellular supremacy in biocomputing. Nat Commun. 2019;10:5250.
6. Kozlov AP. Evolution by Tumor Neofunctionalization: The Role of Tumors in the
