BioSystems 224 (2023) 104839

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Visualizing genomic data: The mixing perspective

William Seitz [a,*], A.D. Kirwan Jr. [b], Krunoslav Brčić-Kostić [c], Petar Tomev Mitrikeski [c,1], P.K. Seitz [d]

[a] Texas A&M University at Galveston, Galveston, TX 77553, United States of America
[b] College of Earth, Ocean and Environment, University of Delaware, Newark, DE, 19716, United States of America
[c] Laboratory of Evolutionary Genetics, Division of Molecular Biology, Ruđer Bošković Institute, Zagreb 10000, Croatia
[d] University of Texas Medical Branch, Galveston, TX 77555, United States of America

ARTICLE INFO

Keywords:
Mixing
Genome models
Randomness
Majorization
Partial order

ABSTRACT

We report on a novel way to visualize genomic data. By considering genome coding sequences, *cds*, as sets of the $N = 61$ non-stop codons, one obtains a partition of the total number of codons in each cds. Partitions exhibit a statistical property known as mixing character which characterizes how mixed the partition is. Mixing characters have been shown mathematically to exhibit a partial order known as majorization (Ruch, 1975). In previous work (Seitz and Kirwan, 2022) we developed an approach that combined mixing and entropy that is visualized as a scatter plot. If we consider all 1,121,505 partitions of 61 codons, this produces a plot we call the theoretical mixing space, *TGMS*. A normalization procedure is developed here and applied to real genomic data to produce the genome mixing signature, *GMS*. Example GMS's of 19 species, including *Homo sapiens*, are shown and discussed.

1. Background

The discovery of the genetic code that connects triplets of nucleotides (codons) to amino acids, and sequences of codons (cds) to proteins and protein subunits, has revolutionized genetics. This has resulted in an explosion of data. Tens of thousands of cds are now available for download via ftp in the RefSeq area of the National Center for Biotechnology Information (NCBI).

The usual statistical analyses of coding sequences typically focus on entropy or randomness. The simplest entropic analysis is the number of each type of nucleotide in the cds. If each nucleotide is present in approximately equal amounts, the cds is close to random. A more discriminatory entropy of a cds is the probability distribution of the codons, p_i, and the associated Shannon entropy:

$$S_{Shannon} = -\Sigma\, p_i\, log(p_i). \tag{1}$$

It is important to note that the Shannon entropy depends solely on the set of p_i and is independent of their order.

Is there more than just randomness to characterize coding sequence distributions? In a study of chirality, Ruch (1975) introduced the notion of mixing character. His approach was radically different from characterizing information by entropy, yet it is completely consistent with the second law of thermodynamics. It also has recently been applied in quantum information (Halpern, 2018; Egloff and Dahlsten, 2015). See Seitz and Kirwan (2022) for the history behind the mixing character concept and the second law of thermodynamics and for applications to large databases based on Shannon entropy.

Another advantage of appealing to mixing character is that it introduced a second function to characterize the statistics of the cds, namely their incomparability, I. Although I is a purely mathematical construct (Trotter, 1992), the notion has considerable appeal in biological sciences as it quantifies the disconnectivity of the underlying data vectors. In the following sections I is defined precisely and illustrated with a simple example.

For genome scale purposes, it is useful to have a statistical model or visualization of the genome to supplement the list of nucleotides in the cds. Data visualization translates information into a visual context, such as a map or graph. Data visualization tools help to visualize characteristics of genomes for comparison to one another and to suggest individual genes of possible interest within a genome. Applications include identifying individual genes associated with disease, identifying groups of genes that generate proteins of related function, and using genome visualizations to compare species/genomes for many biological research purposes.

There are numerous ways to visualize genomes. These visualizations have been used for different purposes such as "(i) analyzing sequence data, both in the context of de novo assembly and of resequencing

* Corresponding author.
E-mail address: seitzw@tamug.edu (W. Seitz).
[1] Present address: Faculty of Philosophy and Religious Studies, University of Zagreb, Jordanovac 110, PO Box 169, 10000 Zagreb, Croatia.

https://doi.org/10.1016/j.biosystems.2023.104839
Received 6 December 2022; Received in revised form 17 January 2023; Accepted 18 January 2023
Available online 20 January 2023

experiments, (ii) browsing annotations and experimental data mapped to a reference genome, and finally, (iii) comparing sequences from different organisms or individuals" (Nielsen and Vidal, 2001). Existing genome visualization tools have been grouped into four types, genome alignment visualization tools, genome assembly visualization tools, genome browsers, and tools to directly compare different genomes with each other for efficient detection of genomic differences (Pavlopoulos, 2015). For oncogenomic data, visualizations such as scatterplots, networks, heatmaps, clusters and combining machine learning and visualization are used (Qu, 2019).

The approach taken here is developed for the purpose of directly comparing different genomes. As indicated in the title, the method utilizes the property of mixing to augment Shannon entropy, so as to produce a scatter plot visualization, which we term the genome mixing signature (GMS).

Section 2 shows the logic behind mixing that leads to the concept of mixing character (Ruch, 1975). Section 3 introduces the concept of incomparability in partial orders and applies it to sets of codons. We show the result for all possible coding sequences that contain 61 codons, which we term the *theoretical genome mixing space (TGMS)*. Section 4 describes real coding sequences of varying lengths. It develops a statistical normalization process to convert or normalize real sequences to those in the TGMS in Section 3. Application of mixing theory to the normalized sequences then gives the genome mixing signatures shown in Figs. 3–7. Section 5 concludes with a discussion of the relation genome mixing signatures to phylogenetic complexity and suggests areas for further study.

The Appendix shows how individual coding sequences in a genome's GMS may be identified along with their proteins so as to explore relationships of individual genes to one another.

2. Mixing

Mixing characterizes sets of objects according to how they can be formed from other sets with the same number of objects. It is a fundamental physical/combinatoric property independent of the specific objects (Seitz and Kirwan, 2022), and has recently been used in physics, particularly with regard to quantum information and entanglement.

We begin by reviewing the logic behind mixing theory (Ruch, 1975). When any set of objects is the union of subsets such that the subsets have no objects in common, one obtains a partitioning of the complete set. For example, consider a set of 6 fruits: 3 apples, 2 pears, 1 orange, 0 plums, 0 grapes, and 0 mangos. The partitioning of this set is the partition [3,2,1,0,0,0]. We refer to this partition as the mixing character. In what follows, it is important to note that so far as mixing is concerned, this set of fruits is equivalent to a set with 0 apples, 1 pear, 2 oranges, 0 plums, 0 grapes, and 3 mangos, though those sets obviously differ in many other ways. For mixing character, the partition is always given in non-increasing order, so both sets mentioned above have mixing characters [3,2,1,0,0,0]. This partitioning is sometimes also represented as [3,2,1]. We will use the latter notation unless specific reference is made to zero elements.

We consider two baskets of fruit, A and B, with partitions [4,1,1] and [3,2,1] and show that B is more mixed than A. Logically, Ruch argued that doing so requires that it is possible to construct type B baskets from collections of type A baskets. Assume basket B contains 3 apples, 2 pears and 1 orange. Mixing the following type A baskets: [4 apples, 1 orange, 1 pear] with one containing [4 pears, 1 apple, 1 orange] and one containing [4 apples, 1 pear, 1 orange] gives a basket with [9 apples, 6 pears, 3 oranges], which is the equivalent of 3 baskets each containing [3 apples, 2 pears, and 1 orange] i.e., three baskets all of type B can be obtained by mixing type A baskets. Hence, baskets of type B are more mixed than baskets of type A. In other words, the partition [3,2,1] is more mixed than [4,1,1]. But if we consider baskets [2,2,2] and [3,1,1,1], it turns out that neither can be formed from combinations of the other, and the partitions [2,2,2] and [3,1,1,1] are incomparably mixed — or simply incomparable. In short, mixing characters are *partially ordered* (Trotter, 1992).

3. Mixing, majorization, and incomparability

Mathematically, the partition of an integer N is defined as the non-increasing list of N integers (including zeros as required) that sum to N. For example, the partitions of $N = 6$ are [6], [5,1], [4,2], [4,1,1], [3,3], [3,2,1], [3,1,1,1], [2,2,2], [2,2,1,1], [2,1,1,1,1] and [1,1,1,1,1,1], where zero elements are not shown. Clearly there is a 1:1 correspondence between mixing characters and integer partitions. While simple in concept, the number of partitions grows rapidly with N, so that when $N = 61$, there are 1,121,505 possible partitions.

In 1903, Muirhead (1903) introduced the majorization partial order for partitions of integers. In 1974 Ruch (1975) proved that majorization is identical to the partial order of mixing characters.

Let $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_N]$ be a partition of an integer N. The criterion of majorization (sometimes called dominance) is that a partition λ majorizes another partition μ if and only if

$$\lambda_1 + \lambda_2 + \cdots + \lambda_m \geq \mu_1 + \mu_2 + \cdots + \mu_m \quad \forall\, m = 1 \text{ to } N. \tag{2}$$

For example, consider two partitions of $N = 6$: $\lambda = [4,1,1,0,0,0]$ and $\mu = [3,2,1,0,0,0]$, discussed earlier. The two sums (termed partial sums) in Eq. (2) are $4,5,6,6,6,6$ and $3,5,6,6,6,6$, and it is easy to see that for each value of m in Eq. (1) the inequality is satisfied. However, for $\lambda = [2,2,2,0,0,0]$ and $\mu = [3,1,1,1,0,0]$, the partial sums are $2,4,6,6,6,6$ and $3,4,5,6,6,6$. The inequality is not satisfied and the partitions are incomparable. (Note, the partial sums include the zero elements.) The incomparability number, I, for each partition is obtained by counting the total number of other partitions that are incomparable to it. This can be done by exhaustive comparisons of all partitions using Eq. (2).

In this paper the objects in the set are codons, and the set is the coding sequence (Seitz and Kirwan, 2022). The codon order we use is given below, where they are read left to right and numbered 1 through 61:

CGA CGC CGG CGT AGA AGG CTA CTC CTG CTT TTA TTG TCA TCC TCG TCT AGC AGT ACA ACC ACG ACT CCA CCC CCG CCT GCA GCC GCG GCT GGA GGC GGG GGT GTA GTC GTG GTT AAA AAG AAC AAT CAA CAG CAC CAT GAA GAG GAC GAT TAC TAT TGC TGT TTC TTT ATA ATC ATT ATG TGG.

The codon ordering choice does not affect the coding sequence distributions because regardless of order, the list of frequencies of the occurrences of codons when arranged in non-increasing order is the same. However to identify individual genes from the NCBI data, the codon order above must be used.

Consider the set of all possible coding sequences with 61 codons. We call this set the *theoretical genome mixing space (TGMS)*. This set contains over a million sequences. For each of these we computed the entropy and the incomparability (Seitz and Kirwan, 2022). Fig. 1, a scatter plot of incomparability, I, vs. Shannon entropy, S, shows the TGMS.

4. Genome mixing signatures

Recall that the majorization partial order applies to sets with the same number of objects. This is clearly not the case for real coding sequences, so the raw mixing characters must be normalized in some manner so that they become members of the theoretical mixing space; in other words, "normalized" to an integer partition of $N = 61$. To accomplish this we chose to normalize each raw mixing character by dividing each element by the total number of codons in the coding sequence and multiplying by 61, the number of codons. This gives, of course, a vector, ω, of real numbers between 0 and 61. This is then mapped to the closest integer partition of 61 as follows.

First, truncate ω to obtain an integer vector η. The sum of the elements in η is obviously less that 61. Next compute the difference

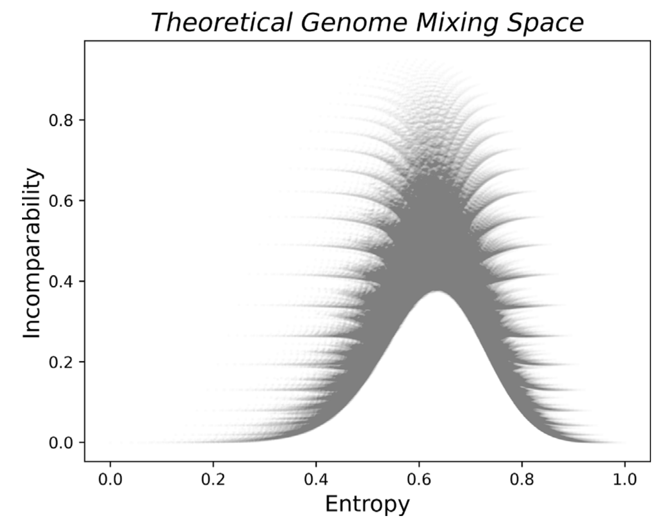

Fig. 1. *Theoretical Genome Mixing Space.*

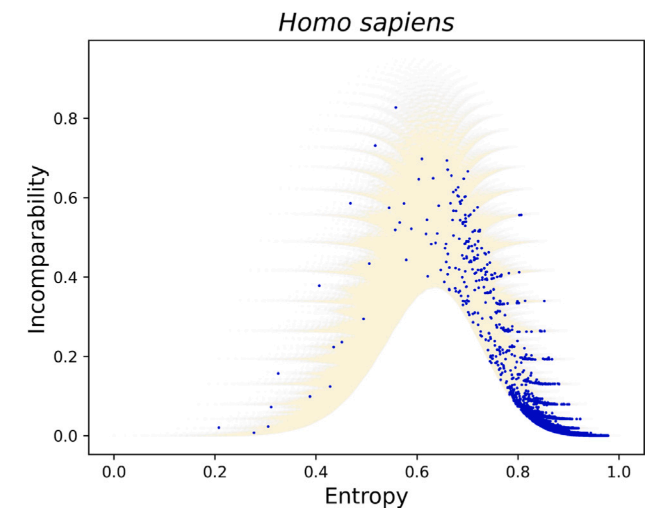

Fig. 2. Humans.

between this vector and the real vector, resulting in a new real vector, v, with elements that lie between 0 and 1. Let λ_i be the ith element of v closest to 1 and increase the corresponding element in η by 1. After setting λ_i to 0, repeat the process until the sum of the elements in η is 61. By choosing 61 as the normalization constant this method obtains mixing characters within the theoretical genome mixing space *TGMS*.

For each normalized coding sequence we compute the entropy, S, and incomparability, I (Seitz and Kirwan, 2018, 2016, 2014) where entropy is normalized to 1 and incomparability is divided by the total number of partitions of $N = 61$. A plot of the set of points (S, I) is the Genome Mixing Signature. We display the GMS of *Homo sapiens* in Fig. 2 superimposed on the theoretical genome mixing space.

Genome mixing signatures for 18 additional species are shown below in Figs. 3–7 where the x-axes are all entropy and the y-axes incomparability. The data sources for these GMS are given in the data section following the references.

The GMS shows two complementary statistical aspects of a genome. One is the entropy or randomness of the cds (x-axis). The other is a statistical measure of how mixed the coding sequence distribution is (y-axis). While gene entropy is well studied, mixing character is a second statistical property that has not been used in comparisons of species genomes.

A GMS considers both the entropy, S, and the incomparability, I of mixing characters. They occupy just a fraction of the theoretical genome mixing space in Fig. 1, since the number of genes in biological species is lower than the number of partitions of $N = 61$. Also, the proportion of occupied mixing space will be even smaller because it is expected that some genes (after normalization) will now have the same mixing character, while their raw mixing characters may still differ. Note that the raw mixing characters can still be obtained from the data, if desired, for more detailed considerations. (See the Data Source section following the references for the NCBI datasets used here.)

5. Discussion

Consider the GMSs of the MERS virus in Fig. 3. From all virus GMS it is clear that they have the smallest number of coding sequences, as was already well known. What is new in the GMS (and for all viruses studied here) is that they appear similar in that they have high entropy and low incomparability. Given that some virus cds have a large number of codons, this need not have been the case. While it is true that many or most virus codon sequences are relatively short, some are not. For example the Middle East virus has 11 coding sequences with codon numbers 586, 313, 139, 76, 68, 62, 48, 36, 32, 24, and 7. They all show entropies greater than 0.8 and incomparabilities below 0.1. Many coding sequences in other (non-virus) species with similar or smaller numbers of codons have lower entropy and higher incomparability. Further studies of additional viruses beyond those in Fig. 3 are required to establish the significance of this difference.

For all other species many coding sequences are less random and more mixed. This is expected since other species are more phylogenetically complex than viruses and have many more coding sequences. This is clearly seen in Figs. 3–7. More importantly, *the mixing signatures of the species studied here are all visually distinct.* Similarities in GMS may indicate biologically relevant relationships between species; this is clearly beyond the scope of the present paper yet remains an area for further work.

For example, a possible area for the application of GMS is in the field of phylogenetics. The phylogenetic relationship between *Alligator mississipiensis* (American alligator) and *Haliaeetus leucocephalus* (bald eagle) illustrates this possibility. These organisms are in different classes (birds and reptiles) yet display similar, though still distinct, GMSs. Genomes of both alligators and birds have the smallest portion of genes with low entropy among the metazoans we have studied; nevertheless, such regularity emphasizes the complex evolutionary relationship between the two (see the GMS for *Alligator mississipiensis* (American alligator) and *Haliaeetus leucocephalus* (bald eagle) in Fig. 7). Members of order Crocodilia are the closest relatives of birds, both being the only surviving representatives of Archosauria (Green et al., 2014; Brusatte et al., 2010), sharing a common ancestor that lived around 240 million years ago. A study of crocodilian genomes led by scientists at UC Santa Cruz (Demuth et al., 2020) revealed an exceptionally slow rate of genome evolution in the crocodilians.

Another interesting application of mixing statistics would be to study transcriptomes, if their sequences can be reliably obtained. This would follow the same procedure as that given here for coding sequences in the NCBI database. A particular area of recent interest is the codon usage bias in co-evolving hosts and viruses (Chen et al., 2020). This is an area for further study where mixing analysis may be of use.

In conclusion, the GMS describes the mixing pattern of gene entropies and incomparabilities independent of chemical details. It thus may provide insights into physical and statistical influences on evolution of individual genes as well as genome evolutionary histories. The relationship between entropy and incomparability was argued in previous work (Seitz and Kirwan, 2018, 2016, 2014) to indicate a relation between incomparability and complexity. However, for genes, sequential complexity is clearly not equal to functional complexity since various gene sequences could have the same incomparabilities and since many biochemically relevant aspects are ignored in a mixing analysis.

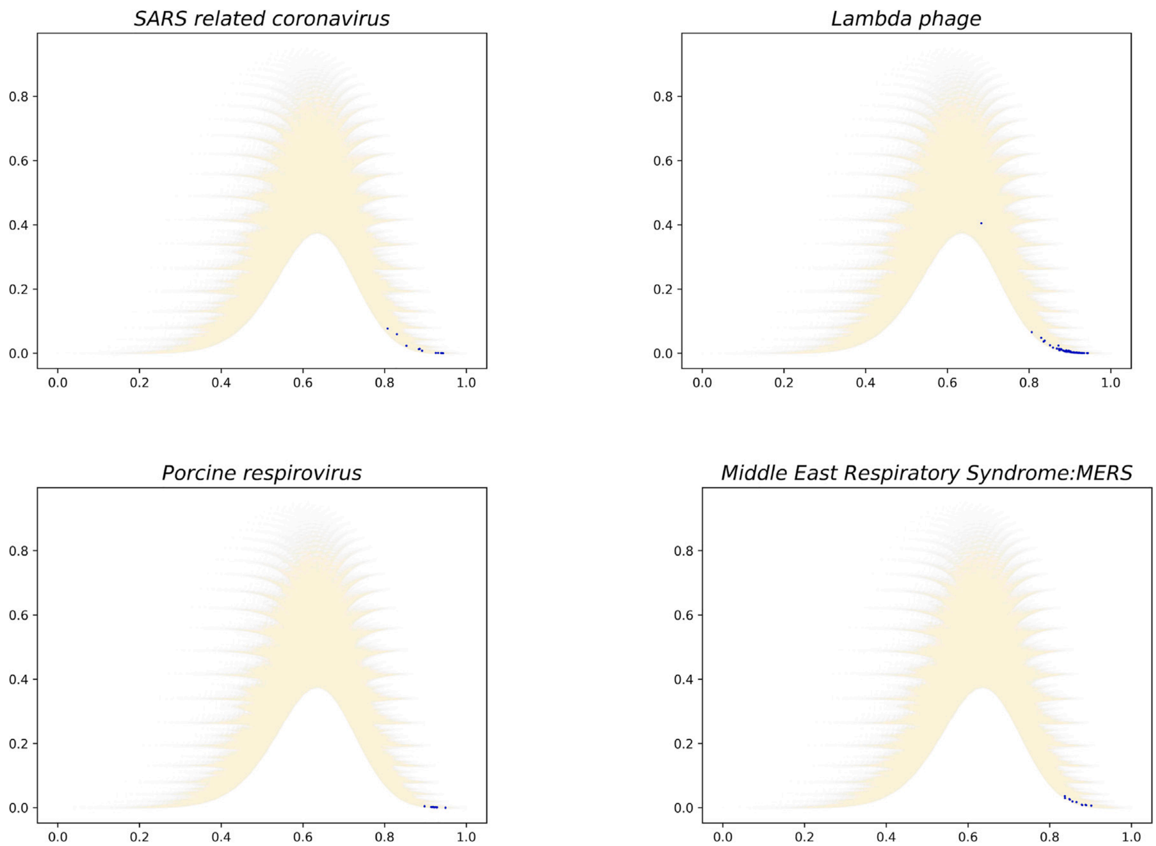

Fig. 3. Viruses.

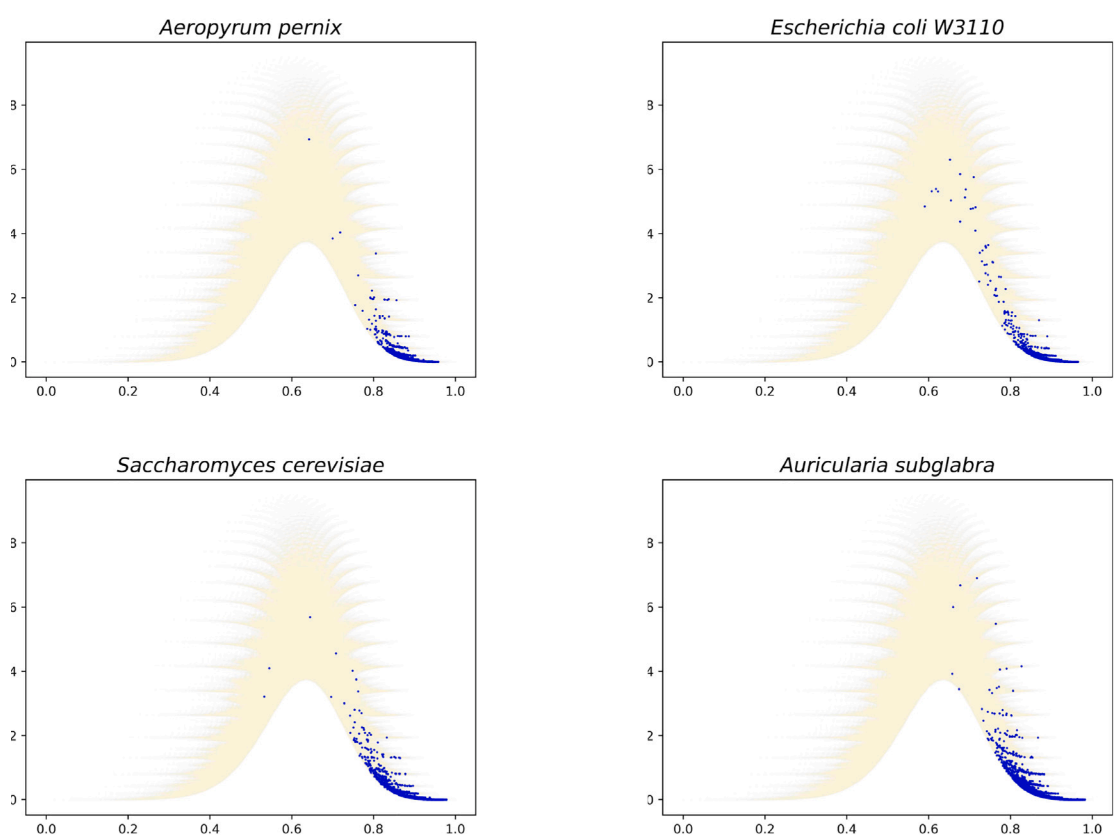

Fig. 4. Archaea, Bacteria, Yeast, and multi-celled Fungi.

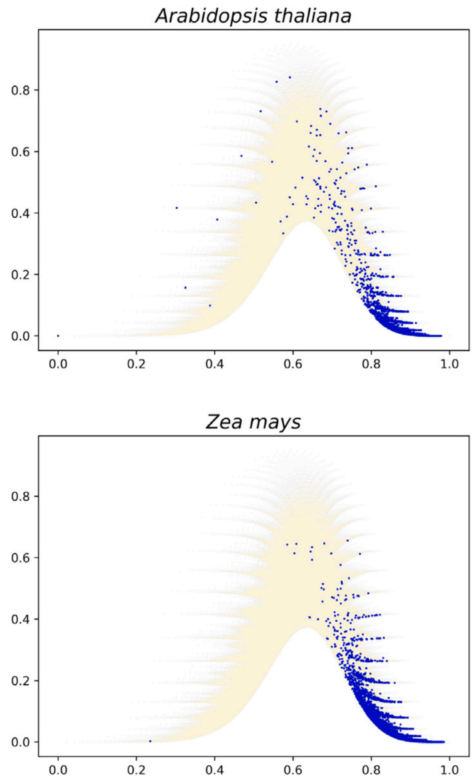

Fig. 5. Plants.

Genome mixing signatures are based on a fundamental statistical property of coding sequences, namely their mixedness. The signatures for different genomes appear to be unique for most species — with viruses a possible exception. The GMS views coding sequences in terms of a statistical scalar quantity quantifying randomness (entropy) and in terms of a vector quantity that characterizes mixing (mixing character). This new visualization of genomes should be of interest to augment visualization tools already in use.

Data source

All data was obtained from the National Center for Biotechnology Information (NCBI) via ftp to the RefSeq area. The file names for all species studied here are listed below.

- *Anopheles gambiae* (*mosquito*) GCF 000005575.2 AgamP3 cds from genomic.fna
- *Eschherichia coli* (*bacteria*) GCF 000010245.2 ASM1024v1 cds from genomic.fna
- *Saccharomyces cerevisiae* (*yeast*) GCF 000146045.2R64 cds from genomic1.fna
- *Arabidopsis thaliana* (*weed*) GCF 000001735.4 TAIR10.1 cds from genomic.fna
- *Caenorhabditis elegans* (*nematode*) GCF 000002985.6 WBcel235 cds from genomic.fna
- *Danio rerio* (*zebrafish*) GCF 00002035.6 GRCz11 CDS from
- *Drosophila melanogaster* (*fruitfly*) GCF 000001215.4 Release 6 plus ISO1 MT cds from genomic.fna
- *Homo sapiens* (*human*) GCF 000001405.39 GRCh38.p13 cds from genomic.fna
- *Mus musculus* (*mouse*) GCF 000001635.26 GRCm38.p6 cds from genomic.fna
- *Zea mays* (*corn*)GCF 000005005.2 B73 RefGen v4 cds from genomic.fna
- *Aeropyrum pernix* (*Archaea*) GCF 000011125.1 ASM1112v1 cds from genomic(1).fna
- *Auricularia subglabbra* (*multicellularfungi*) GCF 000265015.1 Auricularia subglabra SS-5 V1.0 cds from genomic.fna
- *Xenopus tropicalis* (*westernclawedfrog*) GCF 000004195.4 UCB Xtro 10.0 cds from genomic.fna
- *Alligator mississipiensis* (*Americanalligator*) GCF 000281125.3 ASM28112v4 cds from genomic.fna

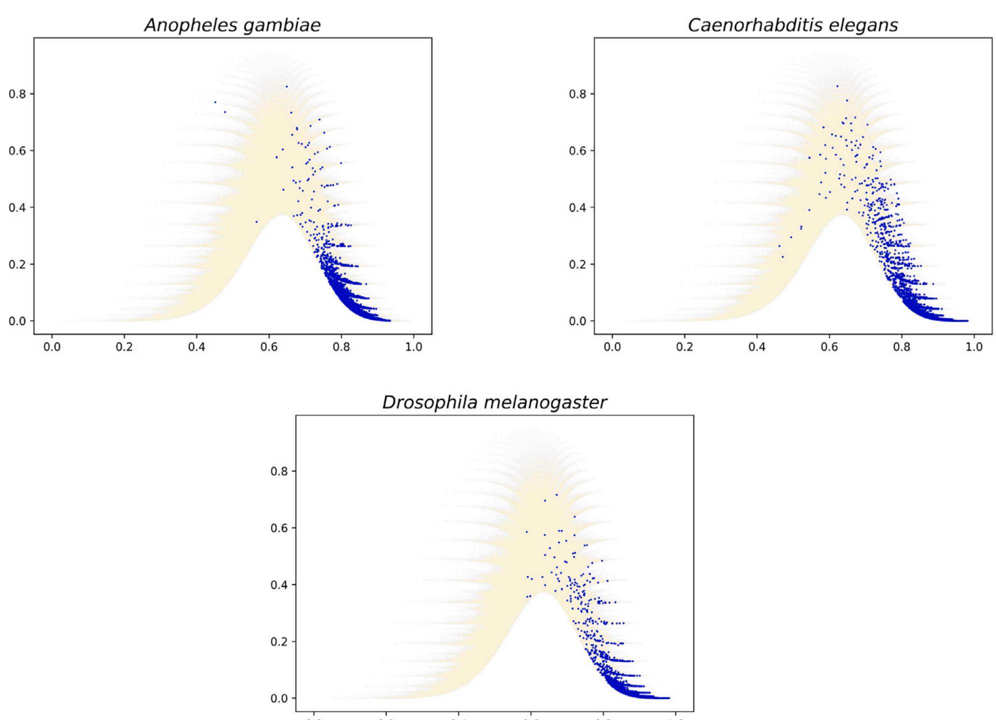

Fig. 6. Invertebrates.

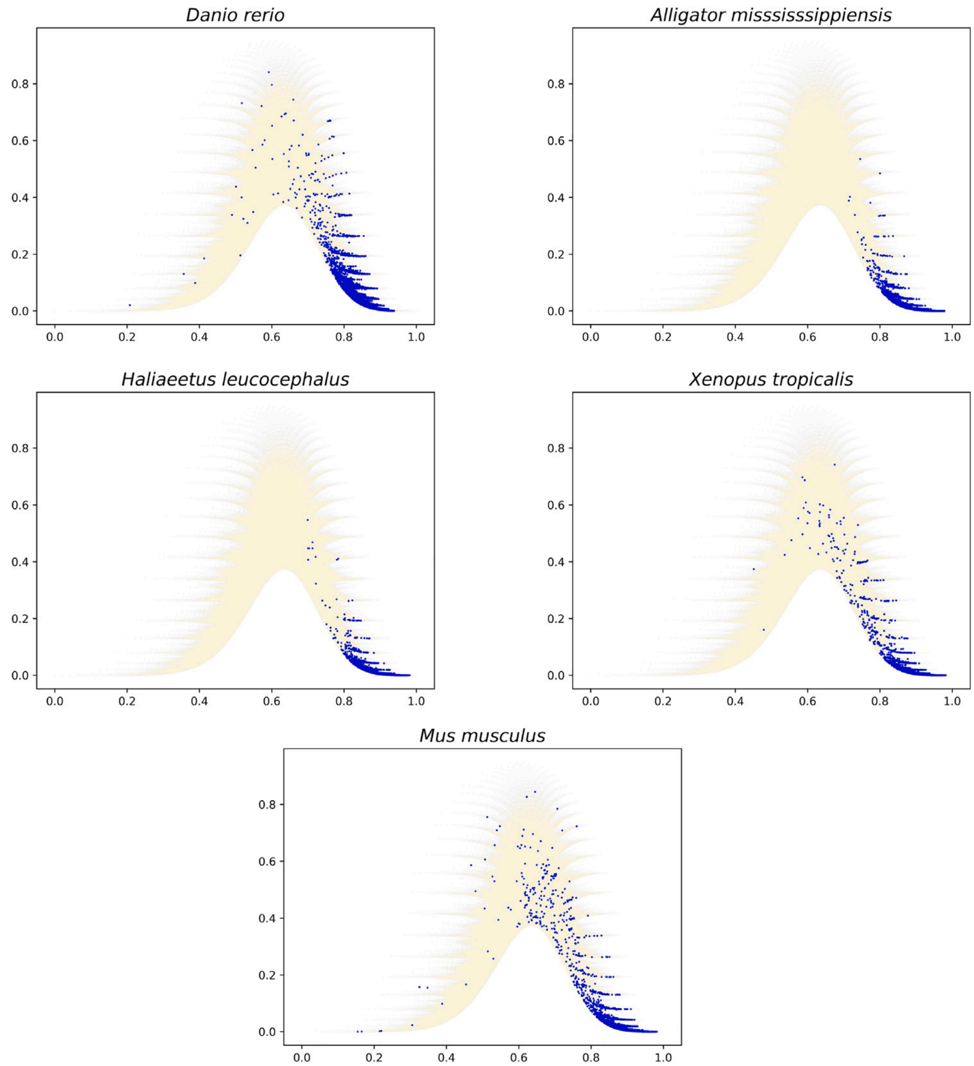

Fig. 7. Vertebrates.

- *Haliaeetus leucoephalus* (*baldeagle*) GCF 000737465.1 Haliaeetus leucocephalus.4.0 cds from genomic.fna
- *Porcine respirovirus* GCF 000925555.1 ViralProj265892 cds from genomic1.fna
- *SARS related coronavirus* GCF 009858895.2A SM985889v3 cds from genomic.fna
- *Middle East Virus, MERS* GCF 000901155.1 ViralProj183710 cds from genomic.fna
- *Lambda phage* GCF 000840245.1vViralProj 14204 cds from genomic.fna

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors thank Karal Gregory for invaluable technical assistance in preparation of the manuscript.

Appendix. Application to single genes

We identified genes in two selected areas of the GMS of two species for study. They are located at different regions of the TGMS. First we consider the genes in *Mus musculus* that have high incomparability and are situated at the top of the inner curve in the mixing space that forms the yellow background in the GMS. Their incomparability values are between 0.37 and 0.422, and the entropy values are between 0.62 and 0.67 (Table 1). There are 18 genes in this region and all of them are connected with keratin (the structural protein of the external part of the epidermis) and hair growth. Hair is unique to mammals. Keratin associated proteins (KRTAPs) that comprise two major groups (high/ultrahigh cysteine and high glycine–tyrosine) are major components of hair and play an essential role in the formation of rigid and resistant hair shafts.

A second example comes from another area of the TGMS, namely very high entropy and very low incomparability.

In the genome of yeast *Saccharomyces cerevisiae*, we analyzed 10 genes with entropy values between 0.969 and 0.978, as well as incomparability values between 1.15916×10^{-5} and 4.72579×10^{-5} (Table 2). It is interesting that 7 genes (*YRF1-1, YRF1-2, YRF1-4, YRF1-5, YRF1-8, YHR218W*, and *YML*133C) are located at subtelomeric regions within the Y' element and are confirmed or putative helicases involved in telomere maintenance via recombination. The rest of

Table 1
Genes in *Mus musculus*.

Entropy	Incomparability	Sequence Number	Name
0.631159454	0.417262524	73128	Gm34733
0.634755753	0.394387901	73137	Keratin-associated
0.635701703	0.416293293	73121	Gm7735
0.635913502	0.392148942	38351	Krtap5-2
0.640166454	0.402695477	38350	Krtap5-3
0.64405205	0.40391618	73115	Gm33798
0.647648349	0.402518928	38354	Krtap5-5
0.647860148	0.407504201	73125	Gm34566
0.648006309	0.381834239	38357	Gm46022
0.648806098	0.421145707	73118	Gm6358
0.650298697	0.383182406	38360	Krtap5-1
0.652044437	0.401277751	73130	Keratin-associated
0.658596899	0.389547974	73133	Gm35174
0.658902185	0.388016999	38353	Keratin-associated
0.659143548	0.40072313	38358	Gm4559
0.659501508	0.378740162	38356	Gm40460
0.660010285	0.382173955	73127	Gm34650
0.663991083	0.370674223	38359	Gm10013

Table 2
Genes in *Saccharomyces cerevisiae*.

Entropy	Incomparability	Sequence Number	Name
0.978422208	1.15916 E-05	3337	AVT3
0.974825909	2.05082 E-05	2049	PUS2
0.971229611	3.29914 E-05	1444	YRF1-1
0.971229611	3.29914 E-05	1731	YRF-2
0.971229611	3.29914 E-05	2688	YHR218 W
0.971229611	3.29914 E-05	4101	YRF1-4
0.971229611	3.29914 E-05	4102	YRF1-5
0.971229611	3.29914 E-05	4103	YML133C
0.971229611	3.29914 E-05	5517	YRF1-8
0.969125911	4.72579 E-05	3590	MMP1

the genes are *AVT3* (involved in amino acid trans membrane export from vacuoles), *MMP1* (involved in S-methylmethionine transmembrane transport) and *PUS2* (involved in mRNA pseudouridine synthesis and in tRNA pseudouridine synthesis). Both AVT3 and MMP1 proteins are yeast amino-acid transporters (Nishida et al., 2016; Popov-Čeleketić et al., 2016) consistent with the observed GMS grouping. Also, as a result of recent reports that PUS2 increases yeast LiCl sensitivity when deleted (Hajikarimlou et al., 2020), we note that AVT5 involved in vascular amino-acids uptake (Chardwiriyapreecha et al., 2010) is also involved in lithium uptake in *Schizosaccharomyces pombe* (Popov-Čeleketić et al., 2016). This is again consistent with the GMS grouping of the three genes reflecting either their functional relatedness or their domain similarities.

References

Brusatte, S.L., Benton, M.J., Desojo, J.B., Langer, M.C., 2010. The higher-level phylogeny of Archosauria (Tetrapoda: Diapsida). J. Syst. Palaeontol. 8, 3–47. http://dx.doi.org/10.1080/14772010903537732.

Chardwiriyapreecha, S., Mukaiyama, H., Sekito, T., Iwaki, T., Takegawa, K., Kakinuma, Y., 2010. Avt5p is required for vacuolar uptake of amino acids in the fission yeast Schizosaccharomyces pombe. FEBS Lett. 584, 2339–2345. http://dx.doi.org/10.1016/j.febslet.2010.04.012.

Chen, F., Wu, P., Deng, S., Zhang, H., Hou, Y., Hu, Z., Zhang, J., Chen, X., Yang, J., 2020. Dissimilation of synonymous codon usage bias in virus-host coevolution due to translational selection. Nat. Ecol. Evol. 4, 589–600.

Demuth, O.E., Rayfield, E.J., Hutchinson, J.R., 2020. 3D hindlimb joint mobility of the stem-archosaur Euparkeria capensis with implications for postural evolution within Archosauria. Sci. Rep. 10, 15357.

Egloff, D., Dahlsten, O., 2015. A measure of majorization emerging from single-shot statistical mechanics. New J. Phys. 17.

Green, R.E., Braun, E.L., Armstrong, J., Earl, D., Ngan, N., Hickey, G., Vandewege, M.W., John, J.A.S., Capella-Gutierrez, S., Castoe, T.A., 2014. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. Science 346 (6215), 1254449. http://dx.doi.org/10.1126/science.1254449.

Hajikarimlou, M., Hunt, K., Kirby, G., Takallou, S., Jagadeesan, S.K., Omidi, K., Hooshyar, M., Burnside, D., Moteshareie, H., Babu, M., Smith, M., Holcik, M., Samanfar, B., Golshani, A., 2020. Lithium chloride sensitivity in yeast and regulation of translation. Int. J. Mol. Sci. 21 (16), 5730. http://dx.doi.org/10.3390/ijms21165730.

Halpern, N.Y., 2018. Beyond heat baths II: framework for generalized thermodynamic resource theories beyond heat baths II: framework for generalized thermodynamic resource theories. J. Phys. A 51 (9).

Muirhead, R.F., 1903. Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters. Proc. Edinb. Math. Soc. 21, 144–157.

Nielsen, M.A., Vidal, G., 2001. Majorization and the interconversion of bipartite states. Quantum Inf. Comput. 1 (1), 76–93.

Nishida, I., Watanabe, D., Tsolmonbaatar, A., Kaino, T., Ohtsu, I., Takagi, H., 2016. Vacuolar amino acid transporters upregulated by exogenous proline and involved in cellular localization of proline in Saccharomyces cerevisiae. J. Gen. Appl. Microbiol. 62, 132–139. http://dx.doi.org/10.2323/jgam.2016.01.005.

Pavlopoulos, G., 2015. Visualizing genome and system biology: technologies, tools, implementation techniques and trends, past, prsent and future. Gigascience 4, 38.

Popov-Čeleketić, D., Bianchi, F., Ruiz, S.J., Meutiawati, F., Poolman, B., 2016. A plasma membrane association module in yeast amino acid transporters. J. Biol. Chem. 291, 16024–16037. http://dx.doi.org/10.1074/jbc.M115.706770.

Qu, Z., 2019. Visual analytics of genomic and cancer data: A systematic review. Cancer Inform. 18.

Ruch, E., 1975. The diagram lattice as structural principle. Theor. Chim. Acta (Berl.) 38, 167–183.

Seitz, W., Kirwan, A., 2022. Mixed-up-ness or entropy. Entropy 24, 1094–1104.

Seitz, W.S., Kirwan, Jr., A.D., 2014. Entropy vs. majorization: What determines complexity? Entropy 16, 3793–3807. http://dx.doi.org/10.3390/e16073793.

Seitz, W.S., Kirwan, Jr., A.D., 2016. Boltzmann complexity: An emergent property of the majorization partial order. Entropy 18, 347.

Seitz, W.S., Kirwan, Jr., A.D., 2018. Incomparability, entropy, and mixing dynamics. Physica A 506, http://dx.doi.org/10.1016/j.physa.2018.05.012.

Trotter, W., 1992. Combinatorics and Partially Ordered Sets. Johns Hopkins University Press.