

Review

Cite this article: Alsufyani D (2024). Evidence of zoonotic pathogens through biophysically induced genomic variance. *Quarterly Reviews of Biophysics*, **57**, e2, 1–8
<https://doi.org/10.1017/S0033583524000039>

Received: 18 August 2023

Revised: 04 January 2024

Accepted: 07 February 2024

Keywords:

computational biophysics; zoonosis; adaptive force; Genodynamics; Adaptation

Email: sufyanid@ksau-hs.edu.sa

Evidence of zoonotic pathogens through biophysically induced genomic variance

Daniah Alsufyani^{1,2,3} 

¹College of Sciences and Health Professions, King Saud bin Abdulaziz University for Health Sciences, King Abdulaziz Medical City, Jeddah, Saudi Arabia; ²King Abdullah International Medical Research Center, Jeddah, Saudi Arabia and ³Ministry of the National Guard - Health Affairs, Jeddah, Saudi Arabia

Abstract

Zoonoses are infectious agents that are transmissible between animals and humans. Up to 60% of known infectious diseases and 75% of emergent diseases are zoonotic. Genomic variation between homeostatic populations provides a novel window into the effect of environmental pathogens on allelic distributions within the populations. Genodynamics is a biophysical approach utilizing developed metrics on biallelic single-nucleotide polymorphisms (SNPs) that can be used to quantify the adaptive influences due to pathogens. A genomic free energy that is minimized when overall population health is optimized describes the influence of environmental agents upon genomic variation. A double-blind exploration of over 100 thousand SNPs searching for smooth functional dependencies upon four zoonotic pathogens carried by four possible hosts amidst populations that live in their ancestral environments has been conducted. Exemplars that infectious agents can have significant adaptive influence on human populations are presented. One discussed SNP is likely associated with both adaptive and innate immune regulation. The adaptive response of another SNP suggests an intriguing connection between zoonoses and human cancers. The adaptive forces of the presented pathogens upon the human genome have been quantified.

Introduction

The availability of population-based genome sequencing has provided an opportunity for exploring the information dynamics of common variants in the human genome. The data presented in common databases (e.g., thousand genome project) are typically presumed to maintain allelic frequencies independent of any subgroups within the data set (Hardy–Weinberg equilibrium) (Hardy, 1908). Migrations and/or admixing of disparate populations likely temporarily take any given population out of Hardy–Weinberg equilibrium. Recent developments in evolutionary biology have modeled the behavior of populations that have been perturbed from homeostasis as genomic evolution toward fitness peaks (Agozzino *et al.*, 2020). Furthermore, applications of statistical physics to this field by others (Sella, 2005) have mirrored unitless free energies describing the evolutionary behaviors of populations. Most such descriptions primarily focus on the evolution of species via mutations rather than adaptation.

For stable populations, we expect that environmental influences affect the frequencies of allelic variations. Adaptation ultimately alters the distribution of these frequencies in a manner that optimizes the overall population health. Zoonoses are particular types of environmental influences involving pathogens that can infect both humans and other species (Chomel, 2014). Therefore, zoonoses provide a window into infections shared via common ancestry.

The degree of variation in any set of variables can be mathematically quantified in terms of the *information content* which typically has no particular set of units associated with it. Most bioinformatic explorations examine unitless measures of allelic variance, and often focus on associations with disease (Uffelmann *et al.*, 2021). Although this is quite useful, such explorations cannot quantify the degree of ‘pressures’ or ‘forces’ that drive human adaptation. On the other hand, genodynamics quantifies the information *dynamics* of common biallelic SNPs as functions of quantifiable environmental agents. This requires the development of a universal *dimensional* unit comparing allelic and genomic ‘energies’, which will be referred to as a genomic energy unit (GEU). For this reason, genodynamics differs from most approaches of standard bioinformatics. These units can then be used to quantify allelic potentials and develop a genomic free energy whose optimization reflects overall population health (Lindesay *et al.*, 2018a).

The original motivation for the development of genodynamics was the quantification of the adaptive responses of allelic distributions due to environmental influences, absent any focus on disease. Calculation of the adaptive *powers* of those influences requires reliable genomic information on the adaptation of populations over substantial periods of time. Furthermore, the response of quasi-homeostatic populations to survivable mutations is *evolution*, which will not be discussed in what follows. The quantification of adaptive *forces* requires that mathematically

smooth functions can describe the variation of allelic frequencies due to environmental differences. Once such functions can be found, the adaptive forces between homeostatic populations due to the environmental differences can be established in terms of the gradients of those functions. Furthermore, genodynamics allows the double-blind explorations of whole population genomes to discover hitherto unknown associations of specific genomic loci with environmental pathogens (Lindesay *et al.*, 2018b).

Several methods and results will be reviewed in some detail. In particular, results relating the adaptive influences due to zoonoses on genomic variation will be emphasized. While performing a double-blind scan of chromosome 3 for adaptive forces on SNPs due to quantified ecological agents, only 0.01% of the SNPs have flagged meaningful relationships. The majority of these flags have involved zoonotic diseases in mammals, indicating a significant influence of the shared pathogens on the optimal distributions of alleles in the populations' genomes. Two such flags will be discussed in some detail. The polymorphism rs1010211 is an intron variant in the gene *TRAF2 and NCK Interacting Kinase* (TNIK) that smoothly adapts to the prevalence of zoonotic viruses. TNIK regulates immune response by activating B cells (which function in the humoral immunity component of the adaptive immune system) (Alsufyani and Lindesay, 2022). Several additional SNPs have indicated particularly interesting possibilities of connections between zoonoses and certain cancers. For instance, the SNP rs16864017 is an intron variant that has flagged a simple mathematical dependency on rodent zoonotic diseases. This variant is located in the gene *tumor protein P63 regulated 1* (TPRG1). Furthermore, current research indicates other possible connections between zoonoses and cancer. Thus, the potential utility of this approach beyond the reviewed results will be explored in the conclusion (Alsufyani and Lindesay, 2023).

Methods and formulation

Genodynamics is a biophysical model that develops adaptive forces and powers, quantifying the information dynamics of the human genome. To explore measures that describe genome variation, the relationship between the information content and entropy is first developed.

Information and entropy

One of the most useful concepts in the physical sciences is that of entropy. To parameterize the thermodynamic descriptions of free energy and entropy (which is related to heat) to statistical physics, Gibbs developed the formula $S = -k_B \sum_a P_a \log P_a$, which describes

the entropy in terms of the statistical probabilities P_a . In this expression, the Boltzmann constant k_B has units of energy per degree Kelvin. However, for general informatics, there are no inherent units associated with the entropy (i.e., $k_B \rightarrow 1$).

For a population that has completely adapted to environmental homeostasis, the distribution of genomic variants within that population optimizes the overall population health. That population will maintain this characteristic allelic distribution to promote its continuing survivability. Genomic variants of particular interest are single-nucleotide polymorphisms (SNPs). The overwhelming majority of SNPs are *biallelic* (i.e., express either of two allelic variants). The maintained order of the SNP variants can be quantified in terms of its entropy. For biallelic SNPs, the specific entropy of a given SNP labeled (S) is defined as:

$$s^{(S)} = - \sum_{a=1}^2 P_a^{(S)} \log_2 P_a^{(S)}, \quad (1)$$

where $s^{(S)}$ is the specific entropy of the SNP, and $P_a^{(S)}$ is the probability (which is related or frequency) of the occurrence of allele a in that population. The entropy is maximum for a completely stochastic SNP distribution (i.e., a population for which both alleles occur with equal probabilities), and it is zero for a homozygous SNP (i.e., only one allele is found among the whole population).

Biological correlations among a vicinity of SNPs manifest within fixed regions of the genome defined as haploblocks. Haploblocks are sets of SNPs with relatively few numbers of haplotypes associated with *individuals*. These haplotypes are transmitted between generations and likewise maintain their frequencies within a population in homeostasis. However, different populations have differing haploblock structures. For correlated SNPs in haploblocks, the haploblock specific entropy $s^{(H)}$ is defined as:

$$s^{(H)} = - \sum_h^{2^{n^{(H)}}} P_h^{(H)} \log_2 P_h^{(H)}, \quad (2)$$

where $n^{(H)}$ is the number of SNPs in haploblock (H) and $P_h^{(H)}$ is the frequency of occurrence of haplotype h in the population. The factor $2^{n^{(H)}}$ represents the maximum *possible* number of haplotypes composed of biallelic SNPs. This identification relates to an SNP entropy as that of a haploblock containing one single SNP. As defined, entropy is an additive variable of the state of the statistical distribution. Therefore, the overall degree of disorder of the whole genome is quantified by the total sum of the specific entropies of both linked and non-linked SNPs to obtain the following:

$$s_{genome} = \left[\sum_H s^{(H)} + \sum_S s^{(S)} \right], \quad (3)$$

Whereas the entropy quantifies the degree of *disorder* within a statistical distribution, the information content $IC = s_{max} - s_{genome}$ quantifies the degree of maintained *order* within the distribution. The entropy s_{max} represents the maximum *possible* whole genome entropy, which for biallelic SNPs is the total number of SNPs in the genome n_{SNPs} . Similarly, the minimum entropy is that of a completely homogeneous population, resulting in a vanishing entropy. Thus, the minimum IC is zero and the maximum is s_{max} . It is convenient to define the normalized information content (NIC) for the whole genome as:

$$NIC_{genome} = \frac{s_{max} - s_{genome}}{s_{max}}. \quad (4)$$

The NIC as defined varies from zero for any maximally disordered distribution to unity for any completely homogeneous distribution. Since it represents a unitless universal measure, the NIC can be used to compare genomes of widely disparate populations (even species), prose written in different languages, and even apples to oranges. For this reason, the NIC can be utilized to interrogate information from different regions of the genome, as well as from differing populations. Furthermore, it gives considerable insight into genomic variations without requiring foreknowledge of biological function or details of genomic history (Lindesay *et al.*, 2012).

Information dynamics

Although an in-depth exploration of the informatics of a statistical distribution can be quite revealing of a multitude of important characteristics, it lacks the potential to explain the *dynamics* of the system being characterized (i.e., how the system changes and responds to stimuli). Biological groups, in particular, present very complicated systems with regard to modeling their information dynamics. For instance, optimizing population health in response to some environmental pathogens (like malaria and its vectors) involves the introduction of particular alleles that can affect the health of an *individual* in either beneficial or detrimental ways. However, sometimes for biallelic variants, one allele is highly advantageous within one extreme environment while the other is highly advantageous in the other extreme.

To quantify the information dynamics of a system, a universal dimensional unit has been introduced as a relative measure of how much of an effect a stimulus has on the system. The dynamic genomic unit should be associated with the degree of genomic variation of the shared variants. For genodynamics, the standard GEU is assigned to environmental agitations that invoke maximum variation of biallelic SNPs (i.e., alleles with equal frequencies) among the population for SNPs that are *not* in linkage disequilibrium.

Once a universal unit of variation has been ascertained, it can be used to compare the relative degree of influence of quantifiable external stimuli on the dynamic system. Genodynamics requires that genomic stimuli be quantified using a smoothly varying well-defined value $\lambda_{stimulus}$ for each stimulus. Once the genomic potential μ_a of allele a (which has units of GEUs) is defined, the adaptive force is then calculated as the gradient down the slope of that genomic stimulus:

$$f_a = -\frac{\partial \mu_a}{\partial \lambda} \tag{5}$$

Genomic potentials will be assigned using guidance provided by energies and potentials defined in physical sciences. The genomic potentials should reflect that populations residing in environments with more vigorous pathogens and stimulants manifest higher degrees of variation when compared to those populations subject to fewer stimulants. When quantifying adaptive forces, it is important that the populations represented by the dynamic variables be in homeostasis with those environments. Although living systems are very far from thermodynamic *equilibrium*, a population in environmental *homeostasis* likewise displays a robust stability under environmental perturbations and fluctuations. In thermodynamics, systems in thermal equilibrium subject to a uniform external agitation (quantified by the temperature T) have their Helmholtz free energy F minimized. This means that this free energy remains unchanged under small fluctuations in quantified stimuli. (i.e., $dF=0$).

Variations in the Helmholtz free energy are typically expressed in terms of temperature T , entropy S , pressure P , volume V , and chemical potentials μ_j associated with population numbers N_j , as given by the following:

$$dF = -SdT - PdV + \sum \mu_j dN_j \tag{6}$$

In thermodynamics, the temperature quantifies a universal set of environmental stimuli that parameterizes the overall degree of agitation of the system. In analogy, the *genomic* free energy of a population in homeostasis with its environment is parameterized by an *environmental potential* T_E that characterizes the overall

degree of agitation and sustenance by the environment. It will be assumed that any ‘pressure’ that the population exerts upon the environment is negligible. For a given population, the genomic free energy is expressed as follows:

$$dF_{genome} = -S_{genome} dT_E + \sum_a \mu_a dN_a + \sum_h \mu_h dN_h, \tag{7}$$

where $\mu^{(S)}$ is the genomic potential of SNP (S) for those SNPs that are not in linkage disequilibrium, and $\mu^{(H)}$ is the genomic potential of haploblock (H). The genomic entropy of the *population* is defined as the size of the population times its specific entropy: $S_{genome} = N_{population} \times s_{genome}$. The dynamic units (GEUs) are carried by the environmental potential and the genomic potentials, and the entropy and population numbers N are dimensionless.

The genomic energy of a homeostatic population should be an *additive* variable of state that depends on the genodynamic variables in Eq. (7). Each SNP potential is the population average of the two allelic potentials in that SNP, $\mu^{(S)} \equiv \sum_a \mu_a^{(S)} P_a^{(S)}$. Likewise, the

haploblock potential $\mu^{(H)}$ is the population average of its haplotype potentials. The population stability condition requires that the size of the population is stable (i.e., $\frac{dN_{population}}{dN_{population}} = 0$). Recognizing that the number of alleles a in the population is the frequency of that allele times the size of the population $N_a^{(S)} = P_a^{(S)} N_{populations}$ Eq. (7) becomes:

$$dF_{genome} = \left[-S_{genome} dT_E + \sum_S \sum_a \mu_a^{(S)} dP_a^{(S)} + \sum_H \sum_h \mu_h^{(H)} dP_h^{(H)} \right] N_{population} + \left[\sum_S \sum_a \mu_a^{(S)} P_a^{(S)} + \sum_H \sum_h \mu_h^{(H)} P_h^{(H)} \right] dN_{population} \tag{8}$$

The population stability condition then implies that the population averaged SNP potentials and block potentials sum to zero over the whole genome $\langle \mu^{(allS)} \rangle + \langle \mu^{(allH)} \rangle = 0$. This condition inherently incorporates Hardy–Weinberg equilibrium (Lindesay *et al.*, 2013; Alsufyani, 2019). The population thus maintains its allelic distribution throughout generations.

For a dynamic population, the allelic potential of a particular SNP should scale with the environmental potential. Differences between allelic potentials should reflect the frequencies of occurrence in an additive manner. These characteristics are incorporated in the dimensionless form:

$$\frac{\mu_{a2}^{(S)} - \mu_{a1}^{(S)}}{T_E} = -\log_2 \frac{P_{a2}^{(S)}}{P_{a1}^{(S)}} \tag{9}$$

This expression clearly vanishes when each probability is $\frac{1}{2}$. As mentioned before, in this case of maximum variation, the allelic potential of each allele is assigned the value $\tilde{\mu} \equiv 1$ GEU. This requires that each potential take the following form:

$$\mu_a^{(S)} = (\tilde{\mu} - T_E) - T_E \log_2 P_a^{(S)} \tag{10}$$

A lower allelic potential is thus assigned to the allele that is more frequent (i.e., more conserved within the population). If the allele is homogenous throughout the population (i.e., $P_a^{(S)} = 1$) then that conserved allele is said to have a fixing potential given by $\mu_{fixing} \equiv \tilde{\mu} - T_E$. Similarly, the haplotype potential has the following form:

$$\mu_h^{(H)} = (\tilde{\mu} - T_E) n^{(H)} - T_E \log_2 P_h^{(H)} \tag{11}$$

The development of potentials with units allows modeling of the dynamics *beyond* just statistical informatics about the population as a whole. Furthermore, one can perform meaningful comparisons of the *effect* of an influence on differing populations with shared SNPs.

Performing the population averages on the genomic potentials from Eqs. (10) and (11), the population stability condition then determines the environmental potential T_E :

$$\mu_{genome} = \langle \mu^{(all S)} \rangle + \langle \mu^{(all H)} \rangle = 0 \rightarrow T_E = \frac{\tilde{\mu} n_{SNPs}}{n_{SNPs} - S_{genome}} = \frac{\tilde{\mu}}{NIC_{genome}}, \quad (12)$$

where n_{SNPs} is the total number of SNPs on the genome. This form is analogous to the thermodynamic temperature of statistical physics in that more homogenous populations have lower environmental potential.

The data associating genomic variance with biological functions (e.g., genome wide association studies) often link phenotypes with individual SNPs. However, alleles that are in linkage disequilibrium mathematically manifest a maintained collective distribution defining a particular haplotype in a haploblock. Such collections of SNPs therefore have decreased entropy, which is reflected through lowered genomic potentials. It is therefore advantageous to distribute a haploblock's potential $\mu^{(H)}$ among its constituent SNPs. The haploblock potential has been distributed using the following criteria (Lindesay *et al.*, 2018a):

- if an allele is homogenous throughout the population, its distributed SNP potential is the same as the fixing potential $\mu_S^{(H)} = \tilde{\mu} - T_E$ (equivalent to a non-linked SNP);
- the distributed SNP potentials are additive and sum to the haploblock potential $\langle \mu^{(H)} \rangle = \sum_S \mu_S^{(H)}$;
- the haploblock potential is distributed among the SNPs proportionate with the degree of allelic variation within each given SNP.

The form of the distributed SNP potential that satisfies these criteria is given by the following:

$$\mu_S^{(H)} = \mu_{fixed} + \left[\mu^{(H)} - n^{(H)} \mu_{fixed} \right] \left[\frac{\bar{P}_s}{\sum_{s'} \bar{P}_{s'}} \right], \quad (13)$$

where \bar{P}_s is the *minor* allele frequency of SNP S' . The haploblock has an increased degree of maintained order when compared to that of the individual SNPs. Therefore, each SNP manifests a binding potential given by the following:

$$\varepsilon_{binding}^{(s)} = \mu_S^{(H)} - \langle \mu^{(s)} \rangle, \quad (14)$$

which is negative, thereby lowering the SNP potential. The definition of a binding potential $\varepsilon_{binding}^{(s)}$ then also allows the distribution of the block potential to the individual alleles within the SNP,

$$\mu_a^{(H)} = \mu_a^{(s)} + \varepsilon_{binding}^{(s)}, \quad (15)$$

which is likewise lowered due to the collective behavior of the alleles.

- Population-based values of environmental potentials:

Finally, the calculation of the various genomic potentials of a given population requires a determination of the overall degree of the environmental agitation of each population characterized by T_E .

Table 1. Environmental potentials of the examined populations

Population	T_E
PEL	1.108
CLM	1.125
FIN	1.125
KHV	1.106
JPT	1.108
TSI	1.122
MSL	1.237
CHB	1.108
IBS	1.110
YRI	1.195

Prior investigations have indicated that the information content of the whole genome is very closely reflected by that of chromosome 3 (Alsufyani, 2019), which has therefore been used to calculate T_E for the various populations. Furthermore, the predictions of genodynamics most accurately describe adaptive forces for those population that have remained in environmental homeostasis for *many* generations. For this reason, data from populations presently residing within their ancestral geographical regions were selected. The populations examined include Peruvian in Lima, Peru (PEL), Colombian in Medellín, Colombia (CLM), Finnish in Finland (FIN), Kinh in Ho Chi Minh City, Vietnam (KHV), Japanese in Tokyo, Japan (JPT), Toscani in Italy (TSI), Mende in Sierra Leone (MSL), Han Chinese in Beijing (CHB), Iberian populations in Spain (IBS), and Yoruba in Ibadan, Nigeria (YRI). The environmental parameters here examined include virus, bacteria, helminth, and protozoa as zoonotic pathogens as well as chiroptera, primates, rodentia, and soricomorpha as zoonotic hosts. The calculated environmental potentials T_E for each population are indicated in Table 1 (Alsufyani, 2019).

Evidence of adaptive forces

Examples of data-based insights into the genomic response to infectious agents transmitted via zoonoses will next be reviewed. Simple functional forms were fitted to the genomic potential versus environmental parameter data. Only quadratic forms for the genomic *potentials* or forms representing simple direct functions of the *frequencies* of allelic occurrence were flagged if the fit satisfied a cut-off criterion. A flagged data set was required to have a root mean squared (RMS) deviation of the data points from the fitted curve of less than 10% of the maximum variation in the potentials. Furthermore, only functional forms that were monotonic in either or both of the alleles were flagged.

Viral zoonotic pathogens

The variant rs1010211 (with common alleles T and C) flagged for an adaptive dependency due to viral pathogen among bats, carnivores, hoofed mammals, moles, primates, rodents, and shrew (Han *et al.*, 2016). The functional behaviors of the SNP potential and that of its alleles are demonstrated in Figure 1.

In the figures, the vertical axes represent the respective genomic potentials in units of GEUs, and the horizontal axes are expressed in

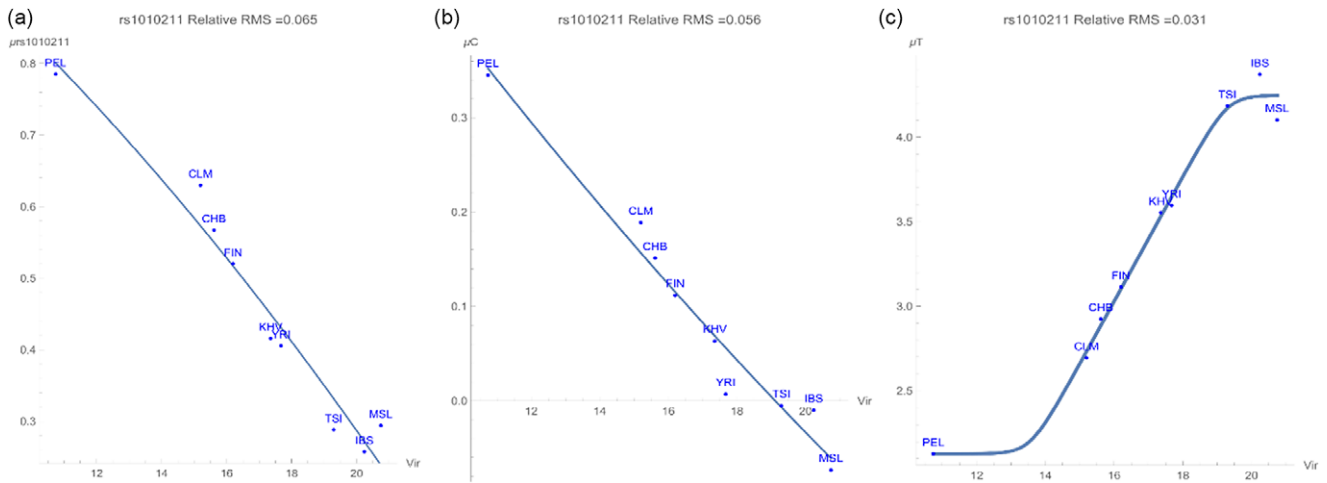


Figure 1. (a) The correlation between the single-nucleotide polymorphism (SNP) rs1010211 and richness pattern of viral pathogens. (b) The correlation between allele C of rs1010211 and richness pattern of viral pathogens. (c) The correlation between allele T of rs1010211 and richness pattern of viral pathogens (Alsufyani and Lindesay, 2022).

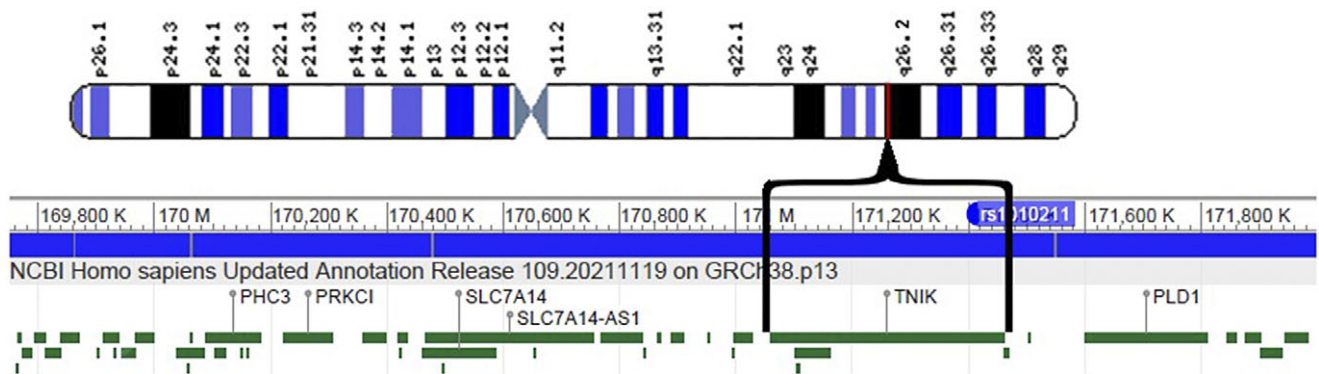


Figure 2. The map of the *TRAF2* and *NCK Interacting Kinase* (TNIK) locus. Chromosome 3 has 199 million base pairs. The gene containing the flagged single-nucleotide polymorphism (SNP), TNIK, extends from 171,058,414 to 171,460,408. The SNP rs1010211 is at locus 171,413,851 (Alsufyani and Lindesay, 2022).

units of viral richness. As defined in the data source (Han *et al.*, 2016),

‘Richness: the number of unique species within a particular geographic area; richness is a count-based metric for quantifying diversity, which contrasts with other metrics, such as functional trait diversity (the different types of traits represented within a geographic area) or genetic diversity.’

The SNP potential plotted in Figure 1a indicates a positive adaptive force of about +0.06 GEU’s/viral zoonosis unit with a relative RMS uncertainty of 0.065. Figure 1b indicates a direct positive adaptive force on the C allele of about 0.04 GEU’s/viral zoonosis unit with relative RMS of 0.056. This allele becomes nearly homogeneous for the populations residing within regions of highest viral zoonoses. In contrast, the T allele plotted in Figure 1c encounters a negative adaptive force of more than −0.2 GEU’s/viral zoonosis unit with relative RMS uncertainty of 0.031. This functional form indicates a direct genomic reaction via the frequency of occurrence of this allele within the population. It should be noted that this SNP is not in linkage disequilibrium for any of the populations considered.

The intron variant rs1010211 lies within the TNIK gene, which is a component of the adaptive immune response. This ancestral

variant is common in zoonotic mammals (*TRAF2* and *NCK interacting kinase*, 2023). TNIK is evolutionarily conserved among amphibians, aves, mammalia, and ray-finned fishes. More than 197 organisms share orthologs retaining the function of human TNIK. This gene also regulates cell division and cell death (Shkoda *et al.*, 2012). TNIK activates B-cells, which function in the humoral immunity component of the *adaptive* immune system (Murphy and Weaver, 2016), producing specialized antibody molecules that then serve as B-cell receptors (Alberts *et al.*, 2002). Recently, TNIK has been found to be a regulator of effector and memory T cell differentiation by inducing a population of undifferentiated memory T cells (Jaeger-Ruckstuhl *et al.*, 2020). Therefore, the fact that this variant responds to viral zoonotic *adaptive* forces is thus confirmed by biophysical quantification.

The TNIK gene map is illustrated in Figure 2.

TNIK is a member of the germinal center kinase (GCK) family (Yu *et al.*, 2014). Germinal centers are transient structures within B lymphocytes that adapt their antibody genes during the immune response to an infection (Natkunam, 2007). These centers play a crucial role in the adaptive humoral immunity component that generates matured B cells, producing effective antibodies against infectious agents. They also play a role in the production of durable

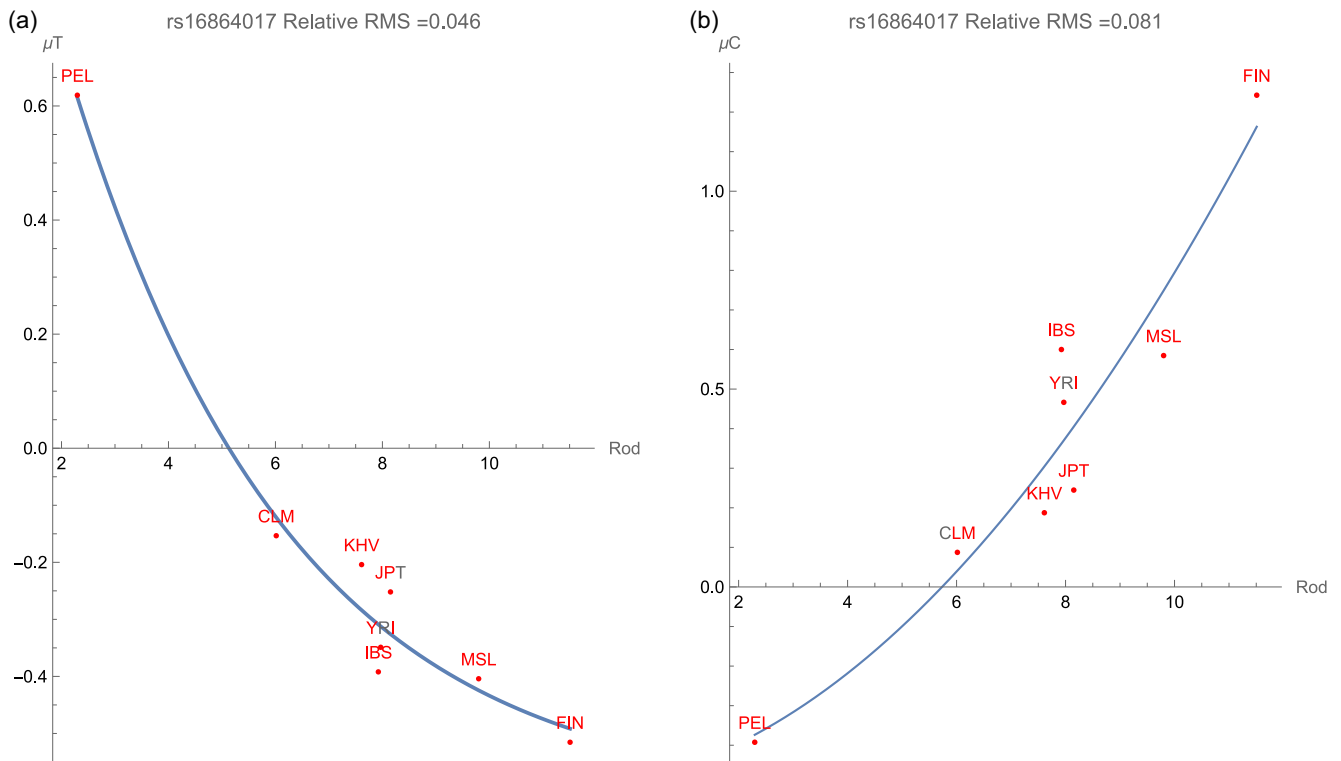


Figure 3. (a) The dependence of allele T of rs16864017 upon richness pattern of rodent pathogens. (b) The dependence of allele C of rs16864017 upon richness pattern of rodent pathogens.

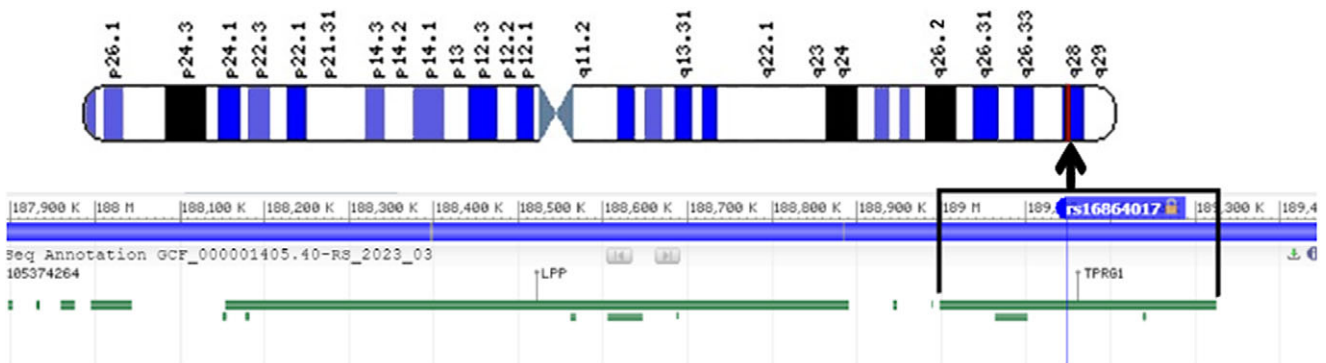


Figure 4. The map of the tumor protein P63 regulated 1 (TPRG1) locus. Chromosome 3 has 199 million base pairs. The gene containing the flagged single-nucleotide polymorphism (SNP), TPRG1, extends from 188997227 to 189325304. The SNP rs16864017 is at locus 189146927.

memory B cells (Yin et al., 2012). It should be noted that GCKs are also involved in innate immune regulation.

Rodents zoonotic pathogens

The intron variant rs16864017 (with common alleles T and C) flagged an adaptive dependency due to rodent zoonoses. The adaptive behaviors of the alleles are demonstrated in Figure 3.

In the figures, the vertical axes represent the respective genomic potentials in units of GEUs, and the horizontal axes are expressed in units of zoonotic richness in rodents.

The T allele in Figure 3a experiences a positive adaptive force of about 0.12 GEUs/rodent zoonosis unit, with relative RMS deviation of 0.046. The collective linked behavior of this allele displays a highly favorable genomic potential for the populations

most exposed to these pathogens. On the other hand, the C allele plotted in Figure 3b is most conserved for the populations that are least exposed to these pathogens, displaying a negative adaptive force of about -0.17 GEUs/rodent zoonosis unit, with relative RMS uncertainty of 0.081. This indicates that the increased presence of rodent zoonosis negatively affects the favorable distribution of the C allele among the population. It should be noted that this SNP is in linkage disequilibrium for all of the populations considered. This suggests a collective biological function shared between this set of SNPs in linkage disequilibrium.

The intron variant rs16864017 lies within the gene TPRG1. The TPRG1 gene map is illustrated in Figure 4.

This gene is differentially expressed in tumor tissues and has been reported to be involved in the regulation of the immune

response (Liu *et al.*, 2019). TPRG1 has orthologs in several chordates, which includes all mammals (and specifically rodents) (*Tumor Protein p63 Regulated 1*, 2023). Knockdown (i.e., temporarily disabling or weakening the expression) of this gene has been found to suppress inflammation in *rats* with cystitis (bladder inflammation), as well as reduce cell proliferation and migration of human primary glandular cells (Hong *et al.*, 2022). It should be noted that several human infections are known to be due to exposure to the urinary and/or respiratory aerosols of rodents (e.g., Hantavirus, *Staphylococcus aureus*, *Streptococcus pneumoniae*) (Williams *et al.*, 2008). TPRG1 has been found to stimulate inflammatory responses (Liu *et al.*, 2019). The expressions of this immune-related gene have also been correlated with early tumor recurrence (Wang *et al.*, 2021). The allele-specific expression (ASE) of TPRG1 has been identified as having tumor-specific expression within the cancer genome atlas as well as in biological evaluation (Th *et al.*, 2020). Furthermore, the ASE in this gene has been found to be specific to human papillomavirus (HPV) positive tumors (Rusan *et al.*, 2015; Kelley *et al.*, 2017; Corces *et al.*, 2018). TPRG1 was also found to be differentially expressed in HPV-associated oropharyngeal squamous cell carcinoma (Th *et al.*, 2020) and breast cancer (Savci-Hejink *et al.*, 2019). It is possible that tumorigenesis is associated with chromosomal rearrangement of TPRG1 in normal lipoma tissues (Wang *et al.*, 2010). Finally, it has been suggested that the antisense strand of this gene TPRG1-AS1 is involved in the suppression of liver cancer growth (Choi *et al.*, 2021).

Conclusion

Populations in environmental homeostasis seem to be accurately described in a manner analogous to physical systems in thermal equilibrium, as described using thermodynamics. In particular, the optimization of overall population health can be used to

Table of abbreviations

Symbols	Definition
$s^{(S)}$	The specific entropy of the SNP.
$s^{(H)}$	The specific entropy of the haploblock
S_{genome}	Overall degree of disorder of the whole genome.
$P_a^{(S)}$	The probability (or frequency) of the occurrence of allele <i>a</i> in that population.
GEUs	Genomic energy unit.
NIC_{genome}	The normalized information content (NIC) for the whole genome.
T_E	Environmental potential.
$\bar{\mu}$	1 GEU.
μ_{fixed}	fixing potential.
μ_a	Genomic potential of allele <i>a</i> .
$\mu^{(S)}$	The genomic potential of SNP (<i>S</i>).
$\mu^{(H)}$	The genomic potential of haploblock (<i>H</i>).
dF_{genome}	The genomic free energy.
$\epsilon_{binding}^{(S)}$	The binding potential.
\bar{P}_S	The <i>minor</i> allele frequency of SNP <i>S'</i> .

characterize the unique distribution of allelic variants within a population satisfying Hardy–Weinberg equilibrium over several generations. In this review, the utility of genodynamics for discovering and quantifying the effects of environmental factors on human adaptation has been explained and exemplified. Zoonoses represent a particular class of environmental pathogens that are shared with other species. A positive adaptive force on the C allele of the SNP variant rs1010211 due to viral zoonoses shared within the same taxonomic class has been quantified. The flagged SNP, which is not in linkage disequilibrium, is an intron variant in the gene that is a component of the adaptive immune response. Furthermore, an interesting adaptive response to rodent zoonotic pathogens on the intron variant rs16864017 is indicative of a possible infectious agent linked to cancer. This SNP is in a gene that is differentially expressed in tumor tissues and is involved in the regulation of the immune response. These results are examples that demonstrate the potential of using genomic information dynamics to understand and describe how environmental factors (including infectious agents) shape the landscape of the human genome.

Acknowledgments. The author is appreciative of the invaluable support and guidance that Dr. James Lindsey extended to this research. His support during the writing process, his commitment to providing insightful feedback, and his meticulous reviews of the manuscript have been instrumental in shaping this work.

Data availability statement. 1000 Genome data is open access. All environmental data used is freely available online.

Code availability statement. Not applicable, formulas can be coded using any programming platform.

Competing interest. The authors declare none.

Resources. Coding: Wolfram Mathematica, version 12.3.1.0.

Genome data: Ensemble (Cunningham *et al.*, 2022).

Environmental data: (Han *et al.*, 2016).

Ethics approval and consent to participate. Data in this research were obtained from open access sources, and it is publicly available.

References

- Agozzino L, Balázi G, Wang J and Dill KA (2020) How do cells adapt? Stories told in landscapes. *Annual Review of Chemical and Biomolecular Engineering* **11**, 155–182.
- Alberts B, Johnson J, Lewis J, Raff M, Roberts K and Walter P (2002) *Molecular Biology of the Cell*. New York: Garland Science.
- Alsufyani D (2019) *Information-Based Analysis of Environmental Factors Adaptive Force and Potential of Sin- Gle Nucleotide Polymorphisms (SNPs) Variants Associated with Human Diseases*. Washington: Howard University.
- Alsufyani D and Lindsey J (2022) Quantification of adaptive forces on SNP rs1010211 due to viral zoonotic pathogens. *Journal of Biological Physics* **48**, 227–236.
- Alsufyani D and Lindsey J (2023) Evidence of cancer-linked rodent zoonoses from biophysical genomic variations. *Scientific Reports* **13**, 13969–13975.
- Choi JH, Kwon SM, Moon SU, Yoon S and Shah M (2021) TPRG1-AS1 induces RBM24 expression and inhibits liver cancer progression by sponging miR-4691-5p and miR-3659. *Liver International* **41**(11), 2788–2800.
- Chomel (2014) *Zoonoses. Reference Module in Biomedical Sciences*.
- Corces M, Granja J, Shams SH, Louie B, Seoane J and Zhou W (2018) The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898.
- Cunningham F, Allen J, Allen J, Alvarez-Jarreta J, Amode M and Armean I (2022) Ensembl 2022. *Nucleic Acids Research* **50**(D1), D988–D995.

- Han BA, Kramer AM and Drake JM** (2016) Global patterns of zoonotic disease in mammals. *Trends in Parasitology* **32**, 565–577.
- Hardy GH** (1908) Mendelian proportions in a mixed population. *Science* **28** (706), 49–50.
- Hong T, Piao S, Sun L, Tao Y and Ke M** (2022) Tumor protein P63 regulated 1 contributes to inflammation and cell proliferation of cystitis glandularis through regulating the NF- κ B/cyclooxygenase-2/prostaglandin E2 axis. *Bosnian Journal of Basic Medical Sciences* **22**(1), 100.
- Jaeger-Ruckstuhl C, Hinterbrandne M, Höpner S, Correnti C, Lüthi U and Friedli O** (2020) NIK signaling imprints CD8+ T cell memory formation early after priming. *Nature Communications* **11**(1), 1632.
- Kelley D, Flam E, Izumchenko E, Danilova L, Wulf H and Guo T** (2017) Integrated analysis of whole-genome ChIP-Seq and RNA-Seq data of primary head and neck tumor samples associates HPV integration sites with open chromatin MarksChIP-Seq/RNA-Seq links HPV integration with open chromatin. *Cancer Research* **77**(23), 6538–6550.
- Lindesay J, Mason TE, Hercules W and Dunston GW** (2013) The foundations of genodynamics: The development of metrics for genomic-environmental interactions. arXiv Preprint, arXiv: 1312.3260.
- Lindesay J, Mason TE, Hercules W and Dunston GW** (2018a) Mathematical modeling the biology of single nucleotide polymorphisms (SNPs) in whole genome adaptation. *Advances in Bioscience and Biotechnology* **9**(10), 520–533.
- Lindesay J, Mason TE, Hercules W and Dunston GW** (2018b). Use of genome information-based potentials to characterize human adaptation. arXiv Preprint, arXiv: 1803.07979.
- Lindesay J, Mason TE, Ricks-Santi L, Hercules W, Kurian P and Dunston GW** (2012) A new biophysical metric for interrogating the information content in human genome sequence variation: Proof of concept. *Journal of Computational Biology and Bioinformatics Research* **4**(2).
- Liu H, Yang Y, Ge Y, Liu J and Zhao Y** (2019) TERC promotes cellular inflammatory response independent of telomerase. *Nucleic Acids Research* **47**(15), 8084–8095.
- Murphy K and Weaver C** (2016) *Janeway's Immunobiology*. New York: Garland Science.
- Natkunam R** (2007) *The Biology of the Germinal Center*. Washington, DC: ASH Education Program Book.
- Rusan M, Li YY and Hammerman PS** (2015) Genomic landscape of human papillomavirus-associated CancersGenomics of HPV-associated cancers. *Clinical Cancer Research* **21**(9), 2009–2019.
- Savci-Heijink CD, Halfwerk H, Koster J, Horlings HM and Van De Vijver M** (2019) A specific gene expression signature for visceral organ metastasis in breast cancer. *BMC Cancer* **19**, 1–8.
- Sella G** (2005) The application of statistical physics to evolutionary biology. *Proceedings of the National Academy of Sciences* **122**, 9541–9546.
- Shkoda A, Town J, Griese J, Romio M, Sarioglu H, Th K and Kieser F** (2012) The germinal center kinase TNK1 is required for canonical NF- κ B and JNK signaling in B-Cells by the EBV oncoprotein LMP1 and the CD40 receptor. *PLoS Biology* **10**(8), e1001376–e1001394.
- Th G, Zambo K, Zamuner F, Ou T, Hopkins CH and Kelley D** (2020) Chromatin structure regulates cancer-specific alternative splicing events in primary HPV-related oropharyngeal squamous cell carcinoma. *Epigenetics* **15**(9), 959–971.
- TRAF2 and NCK interacting kinase** (2023) National Library of Medicine. <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=23043>.
- Tumor Protein p63 Regulated 1** (2023) National Library of Medicine.
- Uffelmann E, Huang Q, Ns M, Vries J, Okada Y, Martin AR and Posthuma HC** (2021) Genome-wide association studies. *Nature Reviews Methods Primers* **59**.
- Wang X, Zamolyi R, Zhang H, Pannain V, Medeiros F and Erickson-Johnson M** (2010) Fusion of HMGA1 to the LPP/TPRG1 intergenic region in a lipoma identified by mapping paraffin-embedded tissues. *Cancer Genetics and Cytogenetics* **96**(1), 64–67.
- Wang Q, Zhou D, Wu F, Liang Q, He Q and Peng M** (2021) Immune microenvironment signatures as biomarkers to predict early recurrence of stage Ia-b lung cancer. *Frontiers in Oncology* **11**, 680287.
- Williams ES, Elizabeth IK and Barker** (2008) *Infectious Diseases of Wild Mammals*. John Wiley & Sons.
- Yin H, Shi Z, Jiao S, Chen C, Wang W, Greene MI and Zhou Z** (2012) Germinal center kinases in immune regulation. *Cellular & Molecular Immunology* **9**(6), 439–445.
- Yu DH, Zhan X, Wang H, Zhang L and Chen H** (2014) The essential role of TNK1 gene amplification in gastric cancer growth. *Oncogenesis* **3**, e89–e89.