

# Information Dynamics of Whole Genome Adaptation

William Hercules<sup>1</sup>, James Lindesay<sup>1</sup>, Tshela E. Mason<sup>2</sup>, Georgia M. Dunston<sup>2,3</sup>

<sup>1</sup>Computational Physics Laboratory, Howard University, Washington DC, USA

<sup>2</sup>National Human Genome Center, Howard University, Washington DC, USA

<sup>3</sup>Department of Microbiology, Howard University, Washington DC, USA

Email: [wmhercules@yahoo.com](mailto:wmhercules@yahoo.com), [jlindesay@howard.edu](mailto:jlindesay@howard.edu), [tshelamason@gmail.com](mailto:tshelamason@gmail.com), [gdunston@howard.edu](mailto:gdunston@howard.edu)

Received 13 August 2014; revised 2 September 2014; accepted 25 September 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

The human genome is a complex, dynamic information system that encodes principles of life and living systems. These principles are incorporated in the structure of human genome sequence variation and are foundational for the continuity of life and human survival. Using first principles of thermodynamics and statistical physics, we have developed analogous “genodynamic tools” for population genomic studies. Characterizing genomic information through the lens of physics has allowed us to develop energy measures for modeling genome-environment interactions. In developing biophysical parameters for genome-environment homeostasis, we found that stable genomic free energy trades off low genomic energy (genomic conservation and increased order) and high genomic entropy (genomic variation) with an environmental potential that drives the variation. In our approach, we assert that common variants are dynamic sites in the genome of a population and that the stability of whole genome adaptation is reflected in the frequencies of maintained diversity in common variants for the population in its environment. In this paper, we address the relativity of whole genome adaptation towards homeostasis. By this we mean that adaptive forces are directly reflected in the frequency distribution of alleles and/or haplotypes of the population relative to its environment, with adaptive forces driving the genome towards homeostasis. The use of genomic energy units as a biophysical metric in DNA sequence variation analyses provides new insights into the foundations of population biology and diversity. Using our biophysical tools, population differences directly reflect the adaptive influences of the environment on populations.

## Keywords

Genomic Information, Population Adaptation, Genomic Homeostasis, Genome-Environment Interactions

---

## 1. Introduction

The completion of the Human Genome Project introduced a new knowledge system for decoding biology and population diversity based on information encoded in the structure of DNA sequence variation. The Biophysics Research and Development Group in the National Human Genome Center at Howard University has applied first principles of thermodynamics and statistical physics in studying common variants (*i.e.*, polymorphisms) as dynamic sites in the genome. This has allowed us to model the genome as a dynamic information system with potentialities for the expression of life in and through all human biology—past, present, and future. These potentialities provide flexibility needed for human survival in varying environments. We submit that on a population level, human survival manifests as maintained diversity associated with whole genome adaptation to the environment. Thus, this paper focuses on genomic information dynamics of whole genome adaptation expressed globally in population diversity. This focus is predicated on the supposition of the human genome as a structured dynamic information system that governs the development, function, and expression of life at all levels of the biological hierarchy, from whole genomes in cells to whole populations.

Our objective in this short population genomics report is to introduce the information dynamics of whole genome adaptation towards homeostasis in population biology. Population *genomics*, the large-scale comparison of DNA sequences of populations focuses on the whole genome as the functional unit. This is in contrast to population *genetics* which focuses on individual genes. In our population genomic studies, the population is defined by the pattern of maintained diversity in the spectrum of allele and/or haplotype frequencies for the population in its environment. This is unlike many conventional studies in population genetics where the population is defined by some subset of the spectrum of allele(s) and/or haplotype(s) most frequent in the population. The constraints of the latter reveal our definition of the population to be more inclusive and comprehensive.

In modeling whole genome adaptation towards homeostasis of the population in its environment, we consider genome-environment interaction as inextricable. Moreover, changes in the distribution of allele and haplotype responses to the environment directly reflect adaptive forces on the population. Thus, adaptive forces drive the genome towards homeostasis. We further note that modeling the information dynamics of whole genome adaptation to environmental stressors does not require detailed knowledge of the molecular mechanisms underlying biology.

## 2. Genomic Information

As a structured, dynamic information system, the information dynamics of the whole genome determines adaptation in a given environment. Genomic information is encoded in patterns of sequence variation. The information content of a genomic distribution is the maintained order reflected in the patterns of sequence variation. This maintained order can be related to the disorder that is parameterized by the entropy  $S$  of the distribution given by

$$S = -\sum_j p_j \log_2 p_j, \quad (1)$$

where  $p_j$  represents the probability of the occurrence of a variant of type  $j$  within the distribution. The degree of order, quantified by the information content is defined using the equation

$$IC = S_{\max} - S \quad (2)$$

where  $IC$  is the information content of the distribution,  $S$  is the previously defined entropy of the distribution and  $S_{\max}$  represents the maximum entropy that results when all possible variations occur with equal likelihood. Using this formula a distribution with zero information content has maximum disorder,  $S = S_{\max}$  whereas a distribution with a normalized information content equal to one contains maximum order,  $S = 0$ . As an information system the human genome must be dynamic, this means that it varies in expression over time and environment. We can therefore model the genome as an information system defined by patterns of sequence variation whose dynamics correlate using genomic energy units. The adaptive forces are determined by how these patterns change as a function of the environmental parameters. The adaptive forces are what drive the changes in the patterns of sequence variation. These forces are directly related to the dynamics of the genomic energies, this therefore ties the dynamics of genomic information to adaptation.

### 3. Population Adaptation

Living systems are inherently adaptive. For our purposes, adaptation is the process by which a population approaches homeostasis in a given environment. Here we will not consider mutations or evolutionary changes in the genome, but rather genomic adaptation involving changes in the frequency distribution of alleles or combinations of alleles (*i.e.* haplotypes). Since the environments in which life exist cover a range of physical, chemical, biological, psychological, and sociological conditions, adaptation is imperative for the continuation of life and human survival. From their original habitat, human populations migrated out of Africa to inhabit all geographic areas spanning the globe.

The movement of human population groups between habitats exposes the migrating groups to different ecological niches which include geography, climate, vegetation, food sources and availability, microorganisms, and predators. That human populations have survived in different ecological environments is evidence of whole genome adaptation as a principle of life and living systems. It is therefore assumed that relatively small adaptive changes occur in populations for them to remain viable in the process of whole genome adaptation towards homeostasis.

Past migrations provide a natural laboratory within which whole genome-environment interactions can be examined, particularly if suitable parameters characterizing the environment can be found. Our research interests in analyzing genome-environment interactions were tremendously augmented by the availability of comprehensive haplotype maps (HapMap) of sequence variation in different continental populations, produced as powerful reference resources to the Human Genome Project's complete sequencing of a reference human genome. The availability of HapMap data on population-associated patterns in sequence variation in different global populations provided the resources to investigate adaptive forces that have influenced patterns of sequence variation in natural populations. Our group has developed biophysical metrics that correlate information dynamics with genomic energy units. We submit that adaptive forces determine how these patterns change as a function of the environmental parameters and are what drive changes in the frequency patterns of sequence variation. Moreover, directly relating these adaptive forces to the dynamics of the genomic energies ties the genomic information dynamics to population adaptation. Ultimately, a population reaches homeostasis characterized by a specific distribution of single nucleotide polymorphisms (SNPs) and SNP haplotypes associated with its adaptation to that environment. Even for a population in homeostasis, continual environmental challenges require maintained diversity.

Population homeostasis involves the maintenance of a state of optimal interaction of its genome with the environment. In contrast with a physical system in thermal equilibrium with its environment, genomic homeostasis maintains a state of *dis-equilibrium*. However, a state of genomic homeostasis can be expressed in terms of state variables parameterizing the information dynamics for the genome in that environment. We define genomic free energy by

$$F_{\text{genome}} = U_{\text{genome}} - T_E S_{\text{genome}} \quad (3)$$

where  $U_{\text{genome}}$  represents a genomic energy whose minimization drives the genomic distribution towards information conservation;  $T_E$  is the environmental potential driving genomic diversity, and  $S_{\text{genome}}$  is the genomic entropy which quantifies the genomic variation. This indicates that a population with given genomic free energy  $F_{\text{genome}}$  has established a balance between the allelic conservation and allelic variation of the population. The parameterization of genomic informatics in terms of genomic free energy makes a statement inherent for living systems that diversity and conservation are inextricable. A population in homeostasis maintains minimal genomic free energy, expressed as

$$dF_{\text{genome}} = 0. \quad (4)$$

Minimal genomic free energy implies that the genomic informatics remains unchanged under small variations in populations (*i.e.* Hardy-Weinberg equilibrium) or small changes in environmental parameters. Once genomic energy units that parameterize the information dynamics of the population have been established in a given environment, these measures can then be used to determine adaptive forces acting upon *individual* alleles and haplotypes. In addition, these measures can be used to determine adaptive forces upon population averaged distribution of alleles. In general, collective adaptive forces drive the genomic distribution towards homeostasis. Using these tools, the information dynamics of whole genome adaptation can be modeled. Modeling the information dynamics of whole genome adaptation to the environment abrogates the need for detailed prior know-

ledge of the underlying biological mechanisms associated with adaptation.

#### 4. Pathogens as Stressors for Modeling Genomic Adaptation

Pathogens are one type of environmental stressors that impact genomic adaptation. Information dynamic tools can be used to model whole genome response to environmental stressors.

An environmental stressor of global significance and particular interest is the plasmodium parasite, the causative agent of malaria. Malaria is considered one of the strongest known adaptive forces for selection in recent history due to the number of selective signatures present in the human genome [1]. Differences between population groups with regard to malaria susceptibility have been of great interest within recent years, and these differences in susceptibility have been linked to differences at the genetic level [2]. Differences have also been found in sympatric populations, where the populations share the same environment but suffer different susceptibility rates and burden of disease [3] [4]. Some of the genetic variation that has arisen in response to malaria infection has been in genes that either directly or indirectly modulates the host immune response, or that are responsible for the regulation of host-parasite interactions [5]. This co-adaptation in both the human and the parasite illustrates the influence the parasite has on the host genome. This has resulted in significant shaping of the host genome. Understanding the information dynamics of whole genome adaptation encoded in human and parasite genomes can offer great insight into the interplay between host-parasite interactions. The genomic information of the parasite may provide insight into the methods utilized by the parasite to produce infection in the host, as well as evasion of the host immune response. Additionally, we expect that the human genome has impacted the parasite's genome and its ability to survive. This clearly illustrates an instance of co-adaptation. Thus host-parasite relationships offer intriguing opportunities to explore mutual information dynamics in co-adaptation.

#### 5. Conclusion

We have used the parameterization of genomic information to derive a genomic energy metric related to genomic free energy. Once genomic energy units that parameterize the dynamics of the population have been established in a given environment, these measures can then be used to determine adaptive forces acting upon individual alleles or haplotypes. In general, collective adaptive forces drive the genome towards homeostasis. Using these tools the information dynamics of whole genome adaptation can be modeled.

#### Acknowledgements

The authors wish to convey appreciation for the continuing support of the National Human Genome Center, and the Computational Physics Laboratory, at Howard University. This research is supported in part by NIH Grant NCRR 2 G12 RR003048 from the RCMI Program, Division of Research Infrastructure. We also acknowledge Zahra Dawson for developing computer programs for calculating biophysical metrics from Haploview raw data.

#### References

- [1] Eid, N.A., Hussein, A.A., Elzein, A.M., Mohamed, H.S., Rockett, K.A., Kwiatkowski, D.P. and Ibrahim, M.E. (2010) Candidate Malaria Susceptibility/Protective SNPs in Hospital and Population-Based Studies: The Effect of Sub-Structuring. *Malaria Journal*, **9**, 119. <http://www.malariajournal.com/content/9/1/119>  
<http://dx.doi.org/10.1186/1475-2875-9-119>
- [2] Kwiatkowski, D.P. (2005) How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria. *The American Journal of Human Genetics*, **77**, 171-190. <http://dx.doi.org/10.1086/432519>
- [3] Modiano, D., Petrarca, V., Sirima, B.S., Nebie, I., Diallo, D., Esposito, F. and Coluzzi, M. (1996) Different Response to *Plasmodium falciparum* Malaria in West African Sympatric Ethnic Groups. *Proceedings of the National Academy of Sciences of the USA*, **93**, 13206-13211. <http://dx.doi.org/10.1073/pnas.93.23.13206>
- [4] Rihet, P., Traore, Y., Abel, L., Aucan, C., Leroux, T.T. and Fumoux, F. (1998) Malaria in Humans: *Plasmodium falciparum* Blood Infection Levels Are Linked to Chromosome 5q31-q33. *The American Journal of Human Genetics*, **63**, 498-505. <http://dx.doi.org/10.1086/301967>
- [5] Driss, A., Hibbert, J.M., Wilson, N.O., Iqbal, S.A., Adamkiewicz, T.V. and Stiles, J.K. (2011) Genetic Polymorphisms Linked to Susceptibility to Malaria. *Malaria Journal*, **10**, 271. <http://www.malariajournal.com/content/10/1/271>  
<http://dx.doi.org/10.1186/1475-2875-10-271>

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either [submit@scirp.org](mailto:submit@scirp.org) or [Online Submission Portal](#).

