Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Universality and predictability in molecular quantitative genetics

Armita Nourmohammad[1,3], Torsten Held[2,3] and Michael Lässig[2]

Molecular traits, such as gene expression levels or protein binding affinities, are increasingly accessible to quantitative measurement by modern high-throughput techniques. Such traits measure molecular functions and, from an evolutionary point of view, are important as targets of natural selection. We review recent developments in evolutionary theory and experiments that are expected to become building blocks of a quantitative genetics of molecular traits. We focus on *universal* evolutionary characteristics: these are largely independent of a trait's genetic basis, which is often at least partially unknown. We show that universal measurements can be used to infer selection on a quantitative trait, which determines its evolutionary mode of conservation or adaptation. Furthermore, universality is closely linked to predictability of trait evolution across lineages. We argue that universal trait statistics extends over a range of cellular scales and opens new avenues of quantitative evolutionary systems biology.

Addresses
[1] Joseph-Henri Laboratories of Physics and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, United States
[2] Institut für Theoretische Physik, Universität zu Köln, Zülpicherstr. 77, 50937, Köln, Germany

Corresponding author: Lässig, Michael (mlaessig@uni-koeln.de)
[3] These authors contributed equally to this work.

Current Opinion in Genetics & Development 2013, **23**:684–693

This review comes from a themed issue on **Genetics of system biology**

Edited by **Shamil Sunyaev** and **Fritz Roth**

For a complete overview see the Issue and the Editorial

0959-437X/$ – see front matter,

http://dx.doi.org/10.1016/j.gde.2013.11.001

Introduction

Quantitative traits are important links between genotypes, organismic functions, and fitness. For some molecular traits, recent sequence data and high-throughput trait measurements have produced quantitative genotype–phenotype maps. Examples include the sequence-dependent binding of transcription factors and histones to DNA, and the formation of RNA secondary structures. For the vast majority of complex traits, however, quantitative genotype–phenotype maps are out of reach. Even comparatively simple molecular traits, such as gene expression levels, depend on a mosaic of *cis*-acting and *trans*-acting sequence loci. We do not know their precise numbers, positions and trait amplitudes, nor relevant evolutionary rates such as the amount of recombination between these loci [1]. This lack of knowledge begs an obvious question: Which evolutionary properties of a quantitative trait are *universal*, that is, independent of these molecular details? In particular, can we formulate natural selection on quantitative traits and their resulting modes of evolution independently of their genetic basis? This article is on universality in molecular evolution. We introduce universality as an emerging statistical property of complex traits, which are encoded by multiple genomic loci. We give examples of experimentally observable universal trait characteristics, and we argue that universality is a key concept for a new quantitative genetics of molecular traits. Three aspects of this concept are discussed in detail. First, universal statistics governs evolutionary modes of conservation and adaptation for quantitative traits, which can be used to infer natural selection that determines these modes. Furthermore, there is a close link between universality and predictability of evolutionary processes. Finally, universality extends to the evolution of higher-level units such as metabolic and regulatory networks, which provides a link between quantitive genetics and systems biology.

Universality in molecular evolution

In a broad sense, universality means that properties of a large system can become independent of details of its constituent parts. This term has been coined in statistical physics, where it refers to macroscopic properties of large systems that are independent of details at the molecular scale [2]. For example, the amount of fluid running through a tube per unit time depends only on the viscosity of the fluid, the diameter of the tube, and the pressure gradient, but not on the detailed chemical composition of the fluid. Thus, rather different fluids have the same flux properties as long as their viscosity is the same. This a strong, experimentally testable statement. It is not always true: if the tube narrows at some point into a nozzle, the fluid becomes turbulent and other things besides viscosity matter. This tells another upfront message: universality is usually not a mathematical identity, but an approximation that is accurate in some cases but not in others.

Universality also arises in evolutionary biology. As in physics, it is a property of systems with a large number of components, and it has strong consequences for experiment and data analysis. In the following, we will discuss a number of examples, and we will pinpoint these components and experimental consequences in each case.

In population genetics, Kimura's celebrated diffusion model for the evolution of allele frequencies is a universal

description [3]. The Kimura model predicts that the frequency distribution of mutant alleles in a large population depends only on the size of the population and the selection coefficients of the alleles, but not on the details of the reproductive process of individuals. Many more detailed models of reproduction, including the Wright–Fisher process [4], the Moran process [5], and branching processes [6], have a common diffusion limit in large populations. Importantly, the universal frequency spectra of the Kimura model are statistical quantities; observing such spectra requires frequency data from a large number of segregating alleles in a population. Hence, the universal spectrum most frequently observed in genomic data is the famous inverse-frequency form for synonymous alleles, which evolve near neutrality. For alleles under selection, universality is often confounded by the heterogeneity of selection coefficients at different genomic loci.

Universality may arise even in Darwinian evolution under strong selection, for example, in rapidly adapting asexual populations. Due to the lack of recombination, competition between simultaneously spreading beneficial mutations leads to complex patterns of rise and fall in their population frequencies. However, if an adaptive process is carried by a large number of segregating alleles, it can be described in a simpler way by a so-called traveling fitness wave [7–12,13•,14•,15,16•]. The speed of this wave, which is also referred to as fitness flux [17], becomes a universal quantity: it depends primarily on rate and average effect of beneficial mutations, but not on the detailed distribution of their selection coefficients [10,12,14•,16•]. In other words, the fitness flux decouples from details of the underlying genomic evolution. More generally, the distribution of fixed mutations becomes insensitive to the details of genomic fitness effects and can be characterized by only a few effective parameters [10,14•,16•]. This feature has also been observed in a microbial evolution experiment under strong selection pressure [18]. Another striking universal feature emerges for 'passenger' mutations carried to fixation by hitchhiking with linked beneficial alleles: their substitution rate becomes independent of their selection coefficients and close to the neutral mutation rate [19,14•,15]. This effect increases the substitution rate of deleterious mutations, which may have significant impact on the adaptive dynamics of pathogens [20] and on cancer progression [21]. Universality has an important consequence for theory: it may allow the construction of models that are simple enough to be solvable, but share their universal properties with more realistic models. In this way, adaptive evolution of asexual populations has recently been mapped onto a solvable stochastic traveling-wave model [13•]. As in physics, this universality has its limitations. For example, if just a few beneficial alleles coexist at a given point in time, the fitness wave starts to stutter and its speed changes [19].

Averaging of allelic contributions is a generic feature in the evolution of quantitative traits. For complex traits, which are encoded by a sufficiently large number of genomic loci, this results in universality. Consider, for example, R.A. Fisher's classic geometric model, which describes the evolution of a trait with d components in a single-peak quadratic landscape of (log, i.e. Malthusian) fitness [22]. In this model, selection favors a unique optimal trait value, but deleterious mutations, which can fix by genetic drift, cause the trait to scatter at some distance from the optimum. This process reaches a selection-mutation-drift equilibrium that depends on the number of trait components, the effective population size, the mutation rate, and the strength of stabilizing selection, but not on details of the genomic loci and their evolution. In this case, the reason underlying universality is compensatory evolution caused by stabilizing selection: individual loci behave in a highly stochastic way, but deleterious changes at one locus tend to be offset by simultaneous or subsequent beneficial changes at other loci. Perhaps the simplest measurable universal quantity is the expected fitness cost or *genetic load* $\mathcal{L} \simeq d/4N$, where N denotes the effective population size. This formula characterizes evolutionary equilibrium for low mutation rates and sufficiently strong stabilizing selection. We derive it in Box 1; variants have been obtained previously in refs. [23–25].

A number of recent studies have treated quantitative traits under stabilizing selection, emphasizing the 'coarse-graining' from genomic alleles to trait variables and the analogy to statistical mechanics [26–31,32•,33•]. Other selection scenarios that have been explored include directional selection [31], adaptation to a moving fitness optimum [33•,34] and apparent selection in the presence of pleitropy [35]. Related statistical methods for the analysis of complex traits are of growing interest for genome-wide association studies [36,1,37,38].

All of these studies cover specific classes of quantitative traits, which is reflected in their assumptions on genome evolution. One group of models applies to *microscopic* quantitative traits, which depend only on a few genomic sites and are generically monomorphic in a population [27–29,43]. For such traits, a population can be approximated as a point in trait space that moves by beneficial and deleterious substitutions; trait diversity and linkage disequilibrium between trait loci are negligible. An example of a microscopic trait is the sequence-dependent binding free energy of a transcription factor to its DNA target sites [44]. Other models treat a complementary class of *macroscopic* or *polygenic* traits, which are encoded by many genomic loci and are always polymorphic. In the spirit of classical quantitative genetics, these models assume fast recombination between the trait loci, which results in complete linkage equilibrium [45,26,30,31,33•]. Interestingly, the evolutionary statistics of a polygenic