Physica D 256–257 (2013) 7–20

Contents lists available at SciVerse ScienceDirect

Physica D

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

Propagation of genetic variation in gene regulatory networks☆

Erik Plahte [a], Arne B. Gjuvsland [a,*], Stig W. Omholt [b,c]

[a] Centre for Integrative Genetics (CIGENE), Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, P.O. Box 5003, N - 1432 Ås, Norway
[b] Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, P.O. Box 5003, N - 1432 Ås, Norway
[c] NTNU Norwegian University of Science and Technology, Department of Mathematical Sciences, N - 7491 Trondheim, Norway

HIGHLIGHTS

- We show that a diploid gene can be modelled as one entity.
- Propagation functions describe how genetic variation propagates through the network.
- Their derivatives can be approximated by observable quantities—and are related to the feedback structure of the system.
- The observable allele interaction value is related to the dominant feedback loop.

ARTICLE INFO

Article history:
Received 24 October 2012
Received in revised form
4 April 2013
Accepted 8 April 2013
Available online 17 April 2013
Communicated by S. Coombes

MSC:
37N25
92C42

Keywords:
Gene regulation
Network
Haploid
Diploid
Genetic variation
Feedback

ABSTRACT

A future quantitative genetics theory should link genetic variation to phenotypic variation in a causally cohesive way based on how genes actually work and interact. We provide a theoretical framework for predicting and understanding the manifestation of genetic variation in haploid and diploid regulatory networks with arbitrary feedback structures and intra-locus and inter-locus functional dependencies. Using results from network and graph theory, we define propagation functions describing how genetic variation in a locus is propagated through the network, and show how their derivatives are related to the network's feedback structure. Similarly, feedback functions describe the effect of genotypic variation of a locus on itself, either directly or mediated by the network. A simple sign rule relates the sign of the derivative of the feedback function of any locus to the feedback loops involving that particular locus. We show that the sign of the phenotypically manifested interaction between alleles at a diploid locus is equal to the sign of the dominant feedback loop involving that particular locus, in accordance with recent results for a single locus system. Our results provide tools by which one can use observable equilibrium concentrations of gene products to disclose structural properties of the network architecture. Our work is a step towards a theory capable of explaining the pleiotropy and epistasis features of genetic variation in complex regulatory networks as functions of regulatory anatomy and functional location of the genetic variation.

1. Introduction

Understanding the *genotype to phenotype map* is essential for a whole range of problems in evolutionary biology, production biology and biomedicine. As gene regulatory networks are the main mediating agents for setting up this map, a theory that can tell us how genetic variation is phenotypically manifested in gene regulatory networks as a function of regulatory anatomy may prove most helpful. Such a theory will be an important contribution to a future quantitative genetics theory linking genes, phenotypes and population level genetic phenomena in causal models based on how genes actually work and interact. More specifically, by being able to describe how the effects of genetic variation propagate in a network one will be able to predict how genetic variation in a gene affects network pathways and processes. In this way one may be able to tie genetic variation in gene networks to a whole range of biological processes that generate high-level phenotypic features. Moreover, at the generic level such a theory can be used in a systematic way to reveal recurrent patterns of how variation is propagated in specific types of regulatory anatomies.

We assume that the network is composed of a set of interacting nodes or loci. Each locus can in principle be regarded as a module by being a functional unit or subsystem of molecular processes whose working may be unknown, but which includes the whole transcriptional and translational machinery that produces the output of the locus [1,2]. The phenotypes of a network are the

* Corresponding author. Tel.: +47 64965292.
E-mail addresses: erik.plahte@umb.no (E. Plahte), arne.gjuvsland@umb.no (A.B. Gjuvsland), stig.omholt@umb.no (S.W. Omholt).

0167-2789/$ – see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physd.2013.04.002

stable equilibrium values of the gene products of all the loci in the network. Each locus is susceptible to genetic variation, and we assume that the genetic variation affects the promoter region of a given gene, but that there is no variation in the coding region of the gene. Many experimental results justify the relevance of this assumption. There are examples of noncoding mutations affecting production rates [3], mRNA processing rates [4,5], the shape of the cis-regulatory input function [6–8], and mRNA decay rates [9–11]. In a recent study of adaptive evolution in threespine sticklebacks, Jones et al. found that in 41% of the genes allelic variation was regulatory, in 42% it was probably regulatory, and in only 17% it was coding [12].

To fully understand the functional properties of a diploid gene it is desirable to model its two alleles as separate quantities. This was first done by Omholt et al. [13] to show how the phenomena of genetic dominance, overdominance, additivity, and epistasis could be seen as generic features of simple diploid gene regulatory networks. This model framework was later used to introduce the so-called *allele interaction* concept [14]. In the present paper, we develop these ideas further by proposing a way by which a diploid gene modelled in this fashion can be represented as a single entity and described by a single ODE for its gene product.

Based on these premises we provide a new vocabulary for analysing how genetic variation is manifested in a wide class of haploid and diploid gene regulatory networks possessing negative and positive feedback loops. We introduce terms to describe how a change in equilibrium value at one locus affects the equilibrium values of all other loci, how to identify the causal chains of loci conveying a genetic signal from one locus to another, and how genetic variation at a particular locus affects the equilibrium value phenotype of the locus itself. In [14] we investigated the relationships between single locus gene action concepts and regulatory network anatomy in small networks. Here we extend the analysis to gene regulatory networks with arbitrary number of loci and complex feedback structures. This extension is highly relevant for understanding *epistasis* and *pleiotropy* in genotype–phenotype maps. Epistasis refers to situations where the effect of a genetic substitution at one locus depends on the genotype at another locus. Pleiotropy describes situations where one gene influences several phenotypes rather than a single one. Since epistasis and pleiotropy are inherent to biological networks, a system-level understanding of these phenomena is needed [15,16].

By this work we contribute to the long and strong tradition originating with the works of René Thomas on relating generic systemic properties to the web of feedback loops [17,18], while at the same time elucidating the link between genetics and systems dynamics. Our results provide further support to the view that nonlinear system dynamics will make up a major part of the core of the mathematical foundation of a future quantitative genetics theory [19,20].

2. Propagation of genetic variation: features shared by haploid and diploid networks

At this stage we are not concerned with the inner workings of each gene due to genetic variation, but assume that the output rate of a locus is a given function of the concentration levels of its regulators, which we assume are one or several gene outputs. Thus in the first part of the paper we deal with characteristics of propagation of genetic variation that are shared by both haploid and diploid networks.

We combine results from linear algebra and graph theory (see e.g. [21]) with gene network ideas to describe how genetic variation in one locus propagates to the other loci in the system in terms of the equilibrium values of the state variables. We introduce the term *propagation function* to describe how a change in equilibrium value of one node affects the equilibrium values of all other nodes, the term *propagation chain* to describe a chain of actions conveying a genetic signal from one node in the network to another, and finally, the term *feedback function* to describe how genetic variation at any particular locus affects the equilibrium value of the locus product itself.

A brief explanation of our notation is found in Appendix A.

2.1. Basic rate equations

We assume the network $\mathcal{N}$ is composed of a set of n loci X_i, $i \in N = \{1, 2, \ldots, n\}$, where $n \geq 2$. The non-negative variable z_i represents the possibly time dependent concentration or amount of the output of X_i and acts as input to other loci in the network or contributes of the network's net output. The dynamics of $\mathcal{N}$ is described by a set of autonomous rate equations E_i for z_i, $i \in N$,

$$\dot{z}_i = f_i(z, a_i) = r_i(z, a_i) - \gamma_i z_i, \tag{1}$$

where $z \in R^n_+$ is the n-component vector with non-negative components z_i, $r_i(z, a_i)$ is differentiable with respect to z in a certain open and convex domain W, and $\gamma_i > 0$ is the relative degradation rate of z_i. The quantity $a = \{a_i\}$, $i \in N$, represents a set of parameters defining the system's genotype, the subset a_i defining the genotype of X_i and comprising quantities like maximum production rate, activation thresholds, affinities of activators and inhibitors, mRNA to protein conversion rate, etc. In many modelling approaches of this type, r_i is a Boolean or Boolean-like functional of sigmoidal functions or piecewise constant functions; see [22] for a review of modelling approaches for gene networks. It should be noted that there could be long and complicated chains of effects incorporated into $r_i(z, a_i)$ [23].

We assume that for each combination of genotypes of the loci X_i in $\mathcal{N}$, the system composed of Eqs. (1) has a single hyperbolic, asymptotically stable and differentiable point-like solution x in W. We show in Section 2.2 that under reasonable assumptions an equilibrium x always exists. If $\mathcal{N}$ has no positive loops, x is unique [24,25]. To avoid having to discuss possible problems related to multistationarity, we invoke the additional assumption that the equilibrium of the system is unique within the domain of phase space of interest even if there are positive loops in the system.

2.2. Propagation functions

A shift in the equilibrium value of some x_k due to a change in parameters specific for X_k will propagate through the network and lead to shifts in other equilibrium values. The propagation follows the network connections, which can be read out from the Jacobian J of Eq. (1) in the stable state x. To the network $\mathcal{N}$ corresponding to Eq. (1) we associate a signed digraph $\mathcal{G}$. To each node or locus X_i is associated a vertex X_i in $\mathcal{G}$. Let $X_j \to X_i$ indicate a direct effect from X_j to X_i if $J_{ij} = \partial r_i(z, a)/\partial z_j \neq 0$ in $z = x$. The effect of X_j on X_i is positive (negative) if the rate of change $\dot{z}_i$ increases (decreases) when z_j increases. For this direct effect there is a corresponding directed arc in $\mathcal{G}$ from X_j to X_i with a sign equal to the sign of J_{ij} associated to it. The sequence of direct effects $X_k \to X_j \to \cdots \to X_l$ is called *a chain from* X_k to X_l if each node in the chain occurs only once [26]. This chain corresponds to a simple path in $\mathcal{G}$ from X_k to X_l. We will use the term *propagation chain*.

The following proposition shows that for each pair $k, l \in N$, where $l \neq k$, there exists a propagation function p_{lk} which determines how the perturbed value of x_l due to a genetic variation in X_k is given in terms of x_k.