BioSystems 139 (2016) 46–54

Contents lists available at ScienceDirect

BioSystems

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

Habitat variability does not generally promote metabolic network modularity in flies and mammals

Kazuhiro Takemoto

Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Iizuka, Fukuoka 820-8502, Japan

ARTICLE INFO

Article history:
Received 30 October 2015
Received in revised form 6 December 2015
Accepted 9 December 2015
Available online 23 December 2015

Keywords:
Network analysis
Habitat variability
Modularity
Metabolic networks
Evolution

ABSTRACT

The evolution of species habitat range is an important topic over a wide range of research fields. In higher organisms, habitat range evolution is generally associated with genetic events such as gene duplication. However, the specific factors that determine habitat variability remain unclear at higher levels of biological organization (e.g., biochemical networks). One widely accepted hypothesis developed from both theoretical and empirical analyses is that habitat variability promotes network modularity; however, this relationship has not yet been directly tested in higher organisms. Therefore, I investigated the relationship between habitat variability and metabolic network modularity using compound and enzymatic networks in flies and mammals. Contrary to expectation, there was no clear positive correlation between habitat variability and network modularity. As an exception, the network modularity increased with habitat variability in the enzymatic networks of flies. However, the observed association was likely an artifact, and the frequency of gene duplication appears to be the main factor contributing to network modularity. These findings raise the question of whether or not there is a general mechanism for habitat range expansion at a higher level (i.e., above the gene scale). This study suggests that the currently widely accepted hypothesis for habitat variability should be reconsidered.

1. Introduction

The evolution of species habitat range is an important topic of scientific inquiry at several scales of biological research, from fundamental biological processes to ecology (Bridle and Vines, 2007; Root et al., 2003; Roy et al., 2009), particularly in the context of predictions related to biodiversity and climate change. Therefore, understanding the factors that determine species habitat use is a relevant topic for advancing these research fields. In particular, it is important to identify the molecular (microscopic) mechanisms that contribute to determining a species habitat range, because the behavior of a species (macroscopic) may result from complex biological systems. For example, some previous studies (Barrett and Schluter, 2008; Kellermann et al., 2009) suggested the importance of genetic variation to the ability to adapt and exploit new environments. Moreover, recent studies reported that gene duplication promotes habitat variability in flies (*Drosophila* species; Makino and Kawata, 2012) and mammals (Tamate et al., 2014). This work was inspired by the proposed importance of gene duplication to increasing biological robustness and evolvability (Wagner, 2008), which are themselves related to habitat variability.

However, it remains unclear how these genetic events affect species habitat variability at a higher level of biological organization. In this context, evaluation of modular organization in biological systems (Hartwell et al., 1999) is useful because it is also generally considered to be related to robustness (Hintze and Adami, 2008) and evolvability (Yang, 2001), despite some opinions to the contrary (Hansen, 2003; Holme, 2011). The evolution of modularity in cellular networks has been specifically intriguing to researchers in the context of network biology (Barabási and Oltvai, 2004; Takemoto, 2012a). In particular, a hypothesis has been proposed that variability in natural habitats promotes network modularity. For example, in a theoretical model, Kashtan and Alon (2005) showed that modular networks spontaneously evolved when a fitness peak determined by the environment changes over time in a manner that preserves the same subgoals but in different permutations. Similarly, Lipson et al. (2002) suggested that changing environments could promote modularity. Hintze and Adami (2008) showed that modularity evolves in biological networks (modeled as metabolic networks) to deal with a multitude of functional goals, with the degree of modularity depending on the extent of environmental variability.

In this context, metabolic networks are particularly interesting because metabolic processes are essential for physiological functions and for maintaining homeostasis in living organisms

E-mail address: takemoto@bio.kyutech.ac.jp

http://dx.doi.org/10.1016/j.biosystems.2015.12.004
0303-2647/

(Takemoto and Oosawa, 2012; Takemoto, 2012a). Metabolic networks also determine the behavior of organisms, such as the space use (Jetz et al., 2004) and feeding rate (Brown et al., 2004) of animals, which may in turn be related to habitat variability. In addition, analyses can be performed using actual empirical data, because metabolic networks are available for a wide diversity of species in databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa et al., 2014) and the Encyclopedia of Metabolic Pathways (MetaCyc; Caspi et al., 2012). In fact, using network analysis, Parter et al. (2007) showed that variability in natural habitats promotes the modularity observed in metabolic networks. This result clearly supports the predictions derived from theoretical models, and several researchers have actively investigated the ecological interactions underlying metabolic networks according to habitat variability (Chave, 2013; Levy and Borenstein, 2012).

Despite this recent attention to this relationship between modularity and habitat viability, more comprehensive examinations are required to resolve some outstanding questions. Indeed, recent studies have cast doubt on this relationship. For example, several alternative theories for explaining the origin and evolution of modularity have been proposed, including the neutral theories of protein (Solé and Valverde, 2008) and metabolic networks (Takemoto, 2012b), connection-cost theory (Clune et al., 2013), and multiplicative-mutation theory (Friedlander et al., 2013). Furthermore, a study conducted in archaea, a type of prokaryote distinct from bacteria, did not find a positive correlation between habitat variability and metabolic network modularity (Takemoto and Borjigin, 2011). Similarly, in bacteria, no positive correlation was observed using the latest version of the metabolic database (Takemoto, 2013; Zhou and Nakhleh, 2012). In short, the observed associations between metabolic network modularity and habitat variability (Parter et al., 2007) may be the result of an artifact due to lack of available data on metabolic reactions. More importantly, the studies conducted thus far are limited to lower organisms such as bacteria and archaea.

Therefore, the aim of this study was to investigate the relationship between habitat variability and metabolic network modularity in higher organisms, including data from flies (Makino and Kawata, 2012) and mammals (Tamate et al., 2014). In addition to the potential effect on habitat variability in promoting network modularity, an association between gene duplication and habitat variability has also been observed. Given that gene duplication also influences the metabolic network structure (Barabási and Oltvai, 2004; Díaz-Mejía et al., 2007; Papp et al., 2004; Takemoto, 2012a), it is reasonable to hypothesize that habitat variability may be linked to not only gene duplication but also metabolic network modularity. To investigate these relationships, data related to habitat variability were collected from the published literature; data were collected only from species for which metabolic network data are also available (see Section 2). Using these data, I evaluated whether habitat variability increases metabolic network modularity and how the association between gene duplication and habitat variability might influence the modularity of the metabolic network.

2. Material and methods

2.1. Collection of data related to habitat variability and fraction of duplicated genes

The data on habitat variability and fraction of duplicated genes in flies and mammals were obtained from Makino and Kawata (2012) and Tamate et al. (2014), respectively. Habitat variability was measured based on the Köppen climate classification in habitat areas for living organisms (see Makino and Kawata, 2012; Tamate et al., 2014 for details). In this study, the Brillouin index was used for measuring habitat variability, because such indices tend to follow a normal distribution for species, which is important for statistical analyses. Previous studies have considered two definitions of habitat variability: the Brillouin index and climate envelope. However, similar conclusions were generally obtained using these two definitions, because of a strong positive correlation between them (Supplementary Tables S1 and S2).

For this analysis, species were selected based on the availability of metabolic network data in the KEGG database (Kanehisa et al., 2014). Finally, data were collected for 11 fly species (*Drosophila* spp.; Supplementary Table S1) and 14 mammalian species (Supplementary Table S2).

2.2. Construction of metabolic networks

The procedure for metabolic network construction is generally the same as that reported previously (Takemoto, 2013). XML files (version 0.7.1) containing the metabolic network data were downloaded from the KEGG database (Kanehisa et al., 2014; ftp://ftp.genome.jp/pub/kegg/xml/kgml/metabolic/organisms/) on August 11, 2015, and two types of metabolic networks were constructed: compound networks and enzymatic networks. As of July 1, 2011, the KEGG FTP site is only available to paid subscribers. Since the use of such data may be desirable to ensure reproducibility, the present dataset on metabolic networks is available upon request.

Compound networks are represented as directed networks, in which the nodes and edges correspond to metabolites and reactions (i.e., substrate–product relationships), respectively. *R* numbers (e.g., R00010) were extracted from the XML files, which indicate metabolic reactions. On the basis of the *R* numbers, substrate–product relationships and reversibility/irreversibility of chemical reactions were identified as carbon traces using the metabolic reaction database (Stelzer et al., 2011). Currency (ubiquitous) metabolites such as H_2O, ATP, and NADH were removed, as described previously (Takemoto et al., 2007).

Enzymatic networks are also represented as directed networks, in which the nodes and edges are metabolic enzymes (reactions) and the presence of interjacent chemical compounds, respectively. In brief, an edge is drawn between 2 enzymes (nodes) if at least 1 product of a reaction catalyzed by an enzyme corresponds to at least 1 substrate of the reaction catalyzed by another enzyme (see Takemoto et al., 2011; Takemoto, 2012a for details). Substrate–product relationships and reversibility/irreversibility of chemical reactions were defined as carbon traces using the metabolic reaction database (Stelzer et al., 2011) in order to avoid the emergence of biologically unsuitable edges (see Takemoto et al., 2011; Takemoto, 2012a for details of the importance of this handling procedure and an example).

The largest (weakly) connected component (giant component) was extracted from each metabolic network to accurately calculate network modularity, and for comparison with the previous study (Parter et al., 2007). In particular, network modularity may be overestimated or underestimated due to isolated components (Parter et al., 2007; Takemoto and Borjigin, 2011; Takemoto, 2013); thus, the use of entire networks was avoided in the present study. However, similar conclusions were also obtained using entire networks.

2.3. Network modularity

The modularity of networks is often measured using the *Q*-value (reviewed by Fortunato, 2010). *Q* is defined as the fraction of edges that lie within, rather than between, modules relative to that expected by chance (see Eqs. (14) and (37) in Fortunato (2010) for the definitions of *Q*s for undirected networks [Q_d] and directed networks [Q_{ud}], respectively).