CORRESPONDENCE BETWEEN STRUCTURAL AND DYNAMIC PROPERTIES IN A BIOCHEMICAL NETWORK

By

JOHN EMMET ANDREWS

A Thesis Submitted to The Honors College

In Partial Fulfillment of the Bachelors degree
With Honors in

Ecology and Evolutionary Biology

THE UNIVERSITY OF ARIZONA

M A Y 2 0 1 6

Approved by:

Dr. Alexander V. Badyaev
Department of Ecology and Evolutionary Biology

ABSTRACT

Phenotypic traits are determined by networks of interactions between genes, proteins and enzymes. Structural positions of regulatory mechanisms in the network can determine variability in its elements. Here we examine the topological properties of controls that regulate element expression in a network. If distinct regulatory controls are located at the beginning and end of the same pathway, then we would expect phenotypic differences to be due to differences in the pathway length of their respective networks. Alternatively, if each pathway has its own control mechanism, then phenotypic differences would be the result of differences in the connection density of pathways present in each network. It is also possible that regulatory mechanisms are independent of the network structure, and if this is the case phenotypic differences would not be related to structural properties of the network. To test these hypotheses we extracted carotenoid compounds from the feathers of 59 female house finches (*Haemorhous mexicanus*) from Arizona and Montana and constructed carotenoid metabolic networks for each individual. We examined whether the highly variable expression of carotenoids in female feathers is caused by differences in the expression of carotenoids as the result of the gain or loss of compounds within the same pathway or among distinct pathways. We find that most of the differences between individuals were caused by differences in which pathways were expressed and not by differences in the elongation of the same pathway. These findings enhance our understanding of biochemical mechanisms that underlie avian carotenoid coloration diversification.

INTRODUCTION

Individuals within a species may express different phenotypes ranging from body size (Sutter et al. 2007) to body form (Segers et al. 2015) to coloration (Lozier et al. 2016). Even genetically identical organisms can express different phenotypes (Storrs & Williams 1968). Phenotypic changes among individuals of the same species can be caused by gene mutations (Khan et al. 2016), or by different environmental conditions such as diet (Greene 1989), temperature (Díaz-Hernández et al. 2015) and oxygen levels (Chi et al. 2006). While phenotypic changes in response to different environmental and genetic perturbations can be observed, how different are these phenotypes from each other? Knowing the magnitude of change between individuals allows an examination of how the specific transition from one phenotype to another occurs, and this has major implications for understanding the roles that phenotypic differences play in the processes of speciation and adaptation.

In order to study the magnitude of difference between phenotypes one can view the phenotype by the deterministic network of the interactions between all of the possible elements such as genes, proteins and enzymes that could underlie the production of the phenotype of a particular trait (Barabasi & Oltvai 2004; Badyaev et al. 2015; Morrison & Badyaev in press). Changes in either the interactions or elements in the network lead to changes in phenotypes, which may create large or small differences depending on what elements are gained or lost (Ma et al. 2007). These differences may conserve elements between different phenotypes, or phenotypes they may share none of the same elements and be completely distinct from each other. Not only is it known which elements are present or absent in a deterministic network, but the structure of the network denotes how these elements are arranged in relation to each other, in terms of which ones are dependent on the presence of others. Some elements may be directly connected by interactions while others may have no affect on each other at all (Barabasi & Oltvai 2004). Knowing the network

structure of the elements that produce a phenotype allows one to determine the specific sequence of changes that lead to differences between phenotypes.

Changes in an individual's expression of a species' network could be caused by mutations of the genes encoding the enzymes or proteins in the network, which could change the actual structure of the potential network that they can use (Vitkup et al. 2006). Alternatively, changes in the expression of a network among individuals of the same species could be due to differences in which of the available elements are used, depending on differences in regulatory mechanisms that switch certain interactions on or off (Morrison and Badyaev in press). In order to investigate these mechanisms it is important to explore the differences in individuals within the same species. This is because it can be assumed that most individuals of the same species generally have the same potential network structure with the same genes encoding for enzymes and proteins in the network, allowing study of the microevolutionary dynamics of phenotypic change and the potential sources of eventual speciation within the network. One can observe the sources of phenotypic changes on the network among individuals that cannot be seen when phenotypic expression is compared between different species, because they have already been lost in some species. This allows one to see the metabolic elements that may potentially be under selection and drive evolutionary differences in network expression.

I hypothesize that one mechanism that drives differences between phenotypes of individuals with access to the same network structure is variation in which of the potential pathways in a species' network are used by an individual (Fig. 1A) due to different regulatory controls acting on each pathway (Morrison and Badyaev in press). These factors could include environments where a certain pathway is more necessary than another, or the cost of production of certain elements within that pathway is higher in one environment than another (Düvel et al. 2010). If this is the case, then the networks of individuals of the same species

should differ more when there are greater differences in the number of pathways they express. Differences in the expression of pathways between networks can be measured by the connectivity of the networks, which is the average number of reactions per compound. The most connected elements in the network are in positions in the network in which several different pathways converge or diverge from the same element, and thus the more pathways one individual uses compared to another will increase the difference in connectivity between the two networks.

Alternatively, phenotypic differences among individuals could be due to variations in the length of a pathway in the network used by both individuals (Fig. 1B). Differences in the pathway length between the networks of individuals in the same species could differ as the result of the necessity for single, intermediate elements to be expressed along a pathway, rather than the end product of a pathway, or to limit the production of unnecessary elements in the network (Brown & Goldstein 2001). If differences in pathway length cause the diversification of network use among individuals, then the networks of individuals should differ more when there are greater differences between the pathway lengths of their networks. When individuals are using the same pathway, I predict that the upstream or starting elements of the pathways in this scenario will be conserved and that the elements located further downstream in the pathway will vary among individuals, with longer pathways being less common across the population.

It may also be possible for groups of compounds to be gained or lost together in the network (Fig. 1C). This may be due to which elements in the network are initially present in an individual. Different parts of the network may be expressed in relation to which precursor elements they have. I predict that networks with the same precursor elements will be more similar to each other, compared to networks whose precursor elements vary.

In order to test my hypotheses I studied the role of network structure in the variable expression of carotenoid compounds in the plumage coloration of female house finches (*Haemorhous mexicanus*). Feather coloration in birds is produced through the use of the global avian carotenoid network, comprised of carotenoid compounds the enzymatic reactions between them, of which house finches utilize specific compounds and enzymatic reactions (Badyaev et al. 2015, Morrison and Badyaev in review). Differences in this use of the global avian carotenoid network result in a diverse range of compound expression and coloration across species of birds (Badyaev et al. 2015, Higginson et al. in press, Morrison and Badyaev in review), but how this occurs on a microevolutionary scale is still unknown If carotenoid network expression is defined by different regulatory mechanisms along a pathway, then the fraction of compounds and reactions that differ between two female networks (metabolic distance) will be greater when there is a larger difference in the number of reactions from the same precursor in a pathway (pathway length) present in each of the networks (Fig. 1A). Compounds located at the beginning of a pathway can be expressed without expressing compounds located further down the same pathway. If carotenoid network expression is defined by different regulatory mechanisms for different pathways, then the metabolic distance between two female networks will be greater when there is a larger difference in the number of reactions per compound (pathway connectivity) between the two networks (Fig. 1B). Distinct pathways originating from the same precursor can be expressed at different times, and the greater the number of reactions associated with a compound, defined as connectivity, the greater chance there is for individuals to use different reactions and different numbers of reactions. Alternatively, if the expression of different pathways is governed by the use of different dietary compounds (compounds only acquired through the diet, from which all other compounds are derived) and not by differences in how the metabolic pathways are used, then the metabolic distance between two female networks will

be greater when there is a greater difference between which dietary compounds are expressed in different individuals, and it will not be related to the difference in the number of metabolically derived compounds between individual networks (Fig. 1C).

To test these hypotheses, I identified the carotenoid compounds in feathers sampled from female house finches from Arizona and Montana over a period of ten years. Networks were then constructed for each individual female based on the compounds identified in their feathers. I then looked at the contributions of differences in dietary compounds, network connectivity and network pathway length to the metabolic differences between female networks. These tests will give insight into the mechanisms that underlie the different phenotypes expressed by individuals within a species.

METHODS

Feather sampling

Rump feathers were sampled from 38 adult female house finches in a free-living individually color-marked study population in southeastern Arizona between 2006 – 2015 and from 21 female house finches in 17 free-living individually color-marked populations in Montana (details in Badyaev and Martin 2000 and Landeen and Badyaev 2012).

Caroteniod extraction and identification

Feather carotenoids were extracted using high-performance liquid chromatography (HPLC). Feathers were trimmed, and the weighed pigmented portions were washed in hexane using Whatman GF/A glass filters and finely ground in 3mL methanol for 10 min at 20Hz using a Retsch MM301 mixer mill (Newtown, PA), equipped with ZrO grinding jars and balls. Carotenoids were extracted using a 0.2µm filter (GHP Arcodisc 13mm Minispike; Pall Life

Sciences, East Hills, NY), and the filtrate was evaporated to dryness under vacuum at 40C and reconstituted in 150µL of HPLC mobile phase (methanol:acetonitrile 50:50, v/v).

Carotenoids were quantified by injecting 50µL of pigment extract into an HPLC System (Shimadzu Corporation, Pleasanton, CA) fitted with a YMC Carotenoid 5.0µm column (250x4.6mm) and guard column (YMC America, Allentown, PA). Analytes were eluted at a constant flow rate of 1.1mL/min using isocratic elution with 42:42:16 (v/v/v) methanol:acetonitrile:dichloromethane for the first 11 min, followed by linear gradient up to 42:23:35 (v/v/v) methanol:acetonitrile:dichloromethane through 21 min, isocratic elution at this condition until 30 min when it returned with step function to the initial isocratic condition at which it was held through 40 min. Carotenoids were detected using a Shimadzu SPD-M10AVP photodiode array detector, and data were collected from 200 to 800nm. Peak areas were integrated at 450 or 470nm depending on the absorbance maximum (λ max) for each compound. Peaks were identified by comparison with the retention times of standard carotenoid compounds (Sigma-Aldrich, St. Louis, MO; Indofine Chemical, Hillsborough, NJ; CaroteNature, Ostermundigen, Switzerland; Santa Cruz Biotechnology, Dallas, TX) and the concentrations of compounds (µg/g) were calculated using calibration curves of these standards.

Network building

The compounds identified in the plumage samples were mapped on the "avian space" of the global carotenoid biosynthesis network (Badyaev et al. 2015). We recorded biochemical pathways that link dietary, intermediate and plumage-expressed compounds for each species. For feather samples that had no known dietary or intermediate compounds (but not both), missing compounds and reactions were assigned based on the mapping of the expressed compounds on the global network and recording all biochemically possible connections (e.g.,

between a known dietary and a known expressed compound or between a known intermediate and a known expressed compound and a possible dietary compound). In the 59 complete networks that were constructed for individual feather samples, 21 compounds and 43 enzymatic reactions occurred at least once.

Network measurements and analyses

I measured pathway length in each network as the longest pathway between a dietary compound and an expressed compound. Network connectivity was calculated using Cytoscape 2.8.2 (Smoot et al. 2011) with NetworkAnalyzer 2.7 (Assenov et al. 2008, Doncheva et al. 2012) and RandomNetworks 1.0 (McSweeney 2008) plug-ins. The difference between each pair of networks with the same dietary compound was calculated for both connectivity and pathway length.

Pairwise metabolic distance (the number of compounds that differ between two networks over the total number of compounds present) was calculated between each pair of individuals. Using these data, I ran partial regressions using SAS 9.4 (SAS Institute, Cary, NC). The first partial regression was between metabolic distance and the differences in the number of dietary compounds and differences in the number of derived (non-dietary) compounds between each pair of networks. In the second partial regression, I analyzed the contributions of the differences between network connectivity and pathway length of pairs of networks with the same dietary compounds to their metabolic difference.

RESULTS

Individuals from Montana had an average of 7.24± 1.05(SE) compounds per network, Arizona individuals had 8.29±0.74, the range was 1 to 19 compounds per network (Fig. 3A). Montana individuals had a pathway length of 1.86±0.30 and Arizona individuals had a

pathway length of 2.16±0.246, the range of pathway length was 1 to 6 (Fig. 3B). Montana individuals had a connectivity of 2.28±0.23 per network and Arizona individuals -- 2.63±0.16, the range of connectivity was 1 to 4.42 (Fig. 3C). The average metabolic distance between pairs of Montana individuals was 0.55±0.02 and the average metabolic distance between Arizona individuals was 4.55±0.01, the range of metabolic distances was 0 to 1 (Fig. 3D).

Partial regressions of pairwise metabolic distances between pairs of female networks and the difference in number of dietary compounds and number of derived compounds in each network indicated no correlation between increases in the number of dietary compounds and the difference in elements present between two metabolic networks (Fig. 4A; b_{ST} = -0.024 t = -1.15, n =1711, P = 0.251). The metabolic difference between networks was greater when there was a greater difference in the number of derived compounds between networks (Fig. 4B; b_{ST} = 0.679, t = 32.24, n = 1711, P <0.001).

Partial regression of pairwise metabolic distances between carotenoid networks with the same dietary compounds and the differences in network connectivity and pathway lengths between pairs of networks indicated that the metabolic difference between networks was greater as the difference in the connectivity of the networks increased (Fig. 4C; b_{ST} = 0.871, t = 34.73, n = 309, P<0.001), while the differences in the pathway lengths between the networks contributed less to the metabolic differences between networks (Fig. 4D; b_{ST} = 0.106, t = 4.23, n=309, P<0.0001).

DISCUSSION

After examining the networks of 59 individual house finches, I found a correspondence between the structural properties of the carotenoid metabolic network and locations of regulatory mechanisms in female house finches. All of the dietary carotenoids were

expressed in each population, but that differences in the number of dietary carotenoids did not contribute to the metabolic distance among networks. However, the difference in the number of derived carotenoids influenced the metabolic differences among networks. Thus, different individuals use different metabolic reactions in the network (Fig. 4B).

To gain insight into how specifically individuals differed in which enzymatic reactions they used in the metabolic network, I examined network structural properties that contributed to metabolic differences between networks that have the same starting dietary compounds. Differences between networks in this case are thus only due to metabolic differences. I found that differences in network connectivity explained more of the metabolic differences between networks (Fig. 4C) than did differences in pathway length (Fig. 4D). This indicates that differences in which compounds individuals express is the result of the use of different numbers of pathways from the same compounds, rather than differences in the length of the pathways they use. The results of this study supports the hypothesis that most of the regulatory differences between individuals are due to variation in how distinct pathways are regulated in the house finch carotenoid metabolic network and that regulatory mechanisms that control the beginning and end of pathways do not seem to be a major source of differences in this system. Thus, the addition of more pathways starting from common compounds shared across individuals appears to be the main mechanism for the diversification of which compounds are expressed in house fiches. The implication of these finding is that the regulation of groups of compounds within a network is a target of natural selection for phenotype divercification. Different regulation of the expression of entire pathways among individuals within a species could drive differences in the adaptation of individuals to different environmental conditions and eventually contribute to speciation (Morrison and Badyaev in press). Future research could examine this possibility as well as examine changes in concentration of compounds in relation to their structural and functional

relationships, based on what we now know about the regulatory control of pathways in the metabolic network.

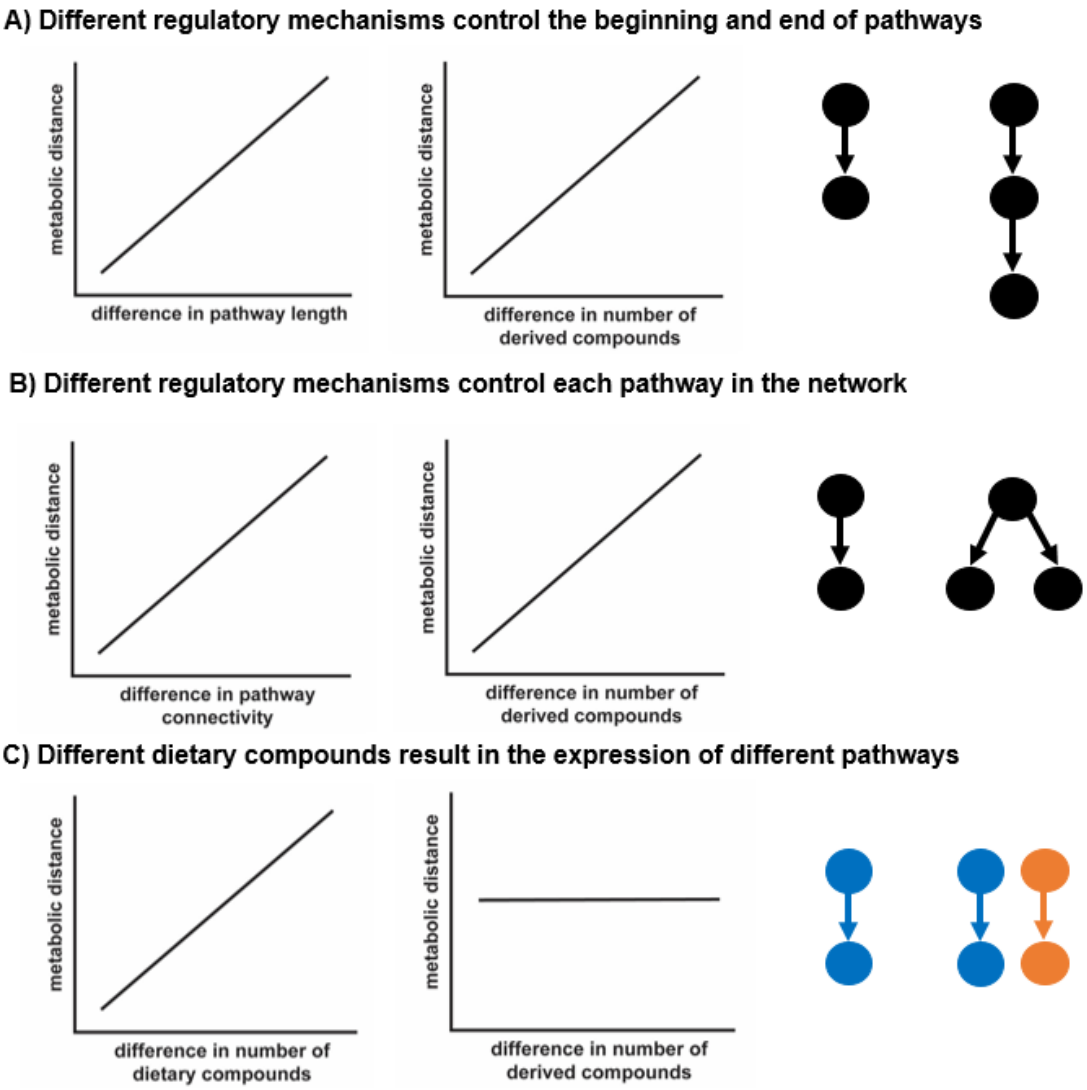

Figure 1: (A) If carotenoid network expression is defined by different regulatory mechanisms along a pathway, then the fraction of compounds and reactions that differ between two female networks (metabolic distance) will be greater when there is a larger difference in the number of reactions from the same precursor in a pathway (pathway length) present in each of the networks. Compounds located at the beginning of a pathway can be expressed without expressing compounds located further down the same pathway. (B) If carotenoid network expression is defined by different regulatory mechanisms for different pathways, then the metabolic distance between two female networks will be greater when there is a larger difference in the number of reactions per compound (pathway connectivity) between the two networks. Distinct pathways originating from the same precursor can be expressed at different times; greater numbers of reactions associated with a compound, defined as connectivity, represent an increase in the number of pathways that are be expressed. (C) If the expression of different pathways is governed by the use of different dietary compounds, then the metabolic distance between two female networks will be greater when there is a greater difference between which dietary compounds are expressed in different individuals, and it will not be related to the difference in the number of metabolically derived compounds between individual networks.

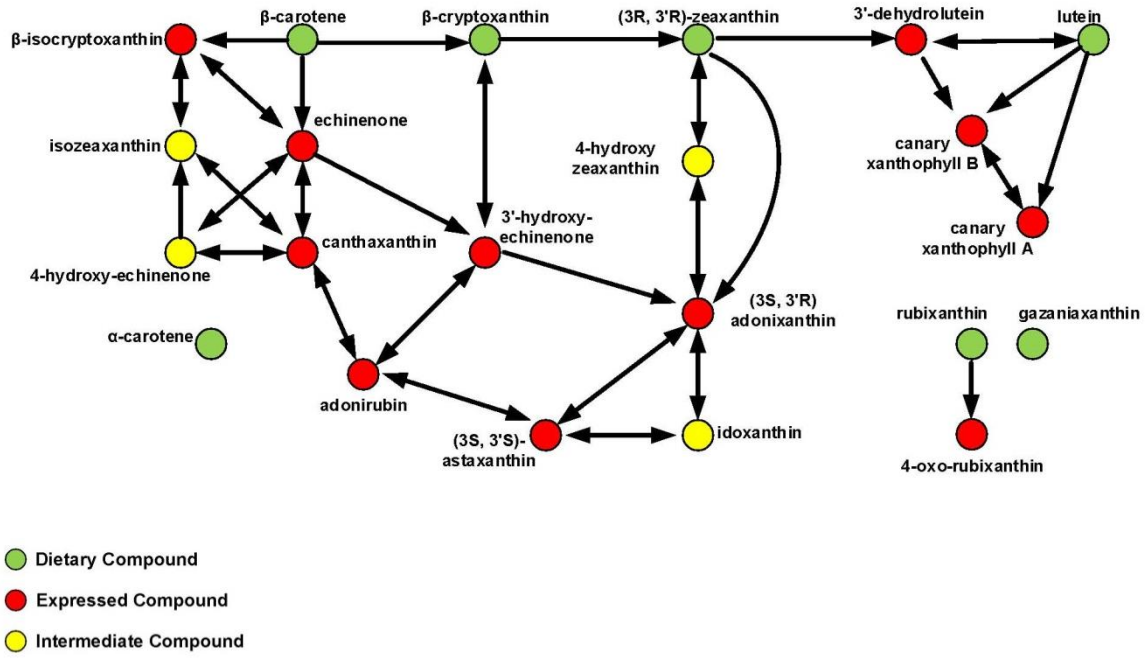

Figure 2. Metabolic network of the 21 carotenoid compounds and 43 reactions identified in female house finches. Birds must consume dietary carotenoids (shown in green) to initiate any enzymatic reactions (arrows) to produce non-expressed intermediate compounds (shown in yellow) and expressed compounds (shown in red).

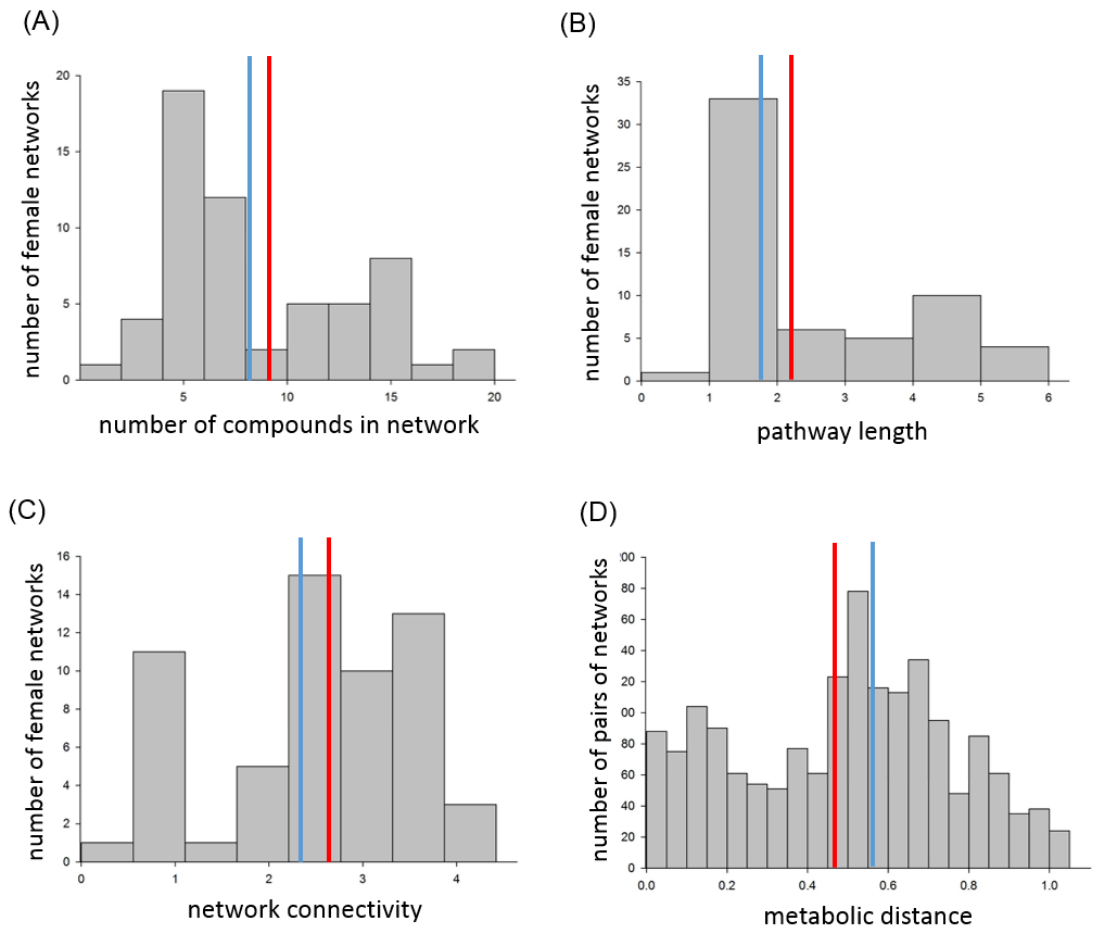

Figure 3. (A) The number of carotenoid compounds in each female network. (B) The pathway length of each female network (largest number of reactions between an expressed compound and a dietary compound). (C) The connectivity (average number of reactions per compound) in each female network. (D) The number of pairs of female networks grouped by metabolic distance (fraction of reactions and compounds that differ between two networks). Lines over each graph indicate the average value for females sampled from the Arizona population (red) and from the Montana population (blue).

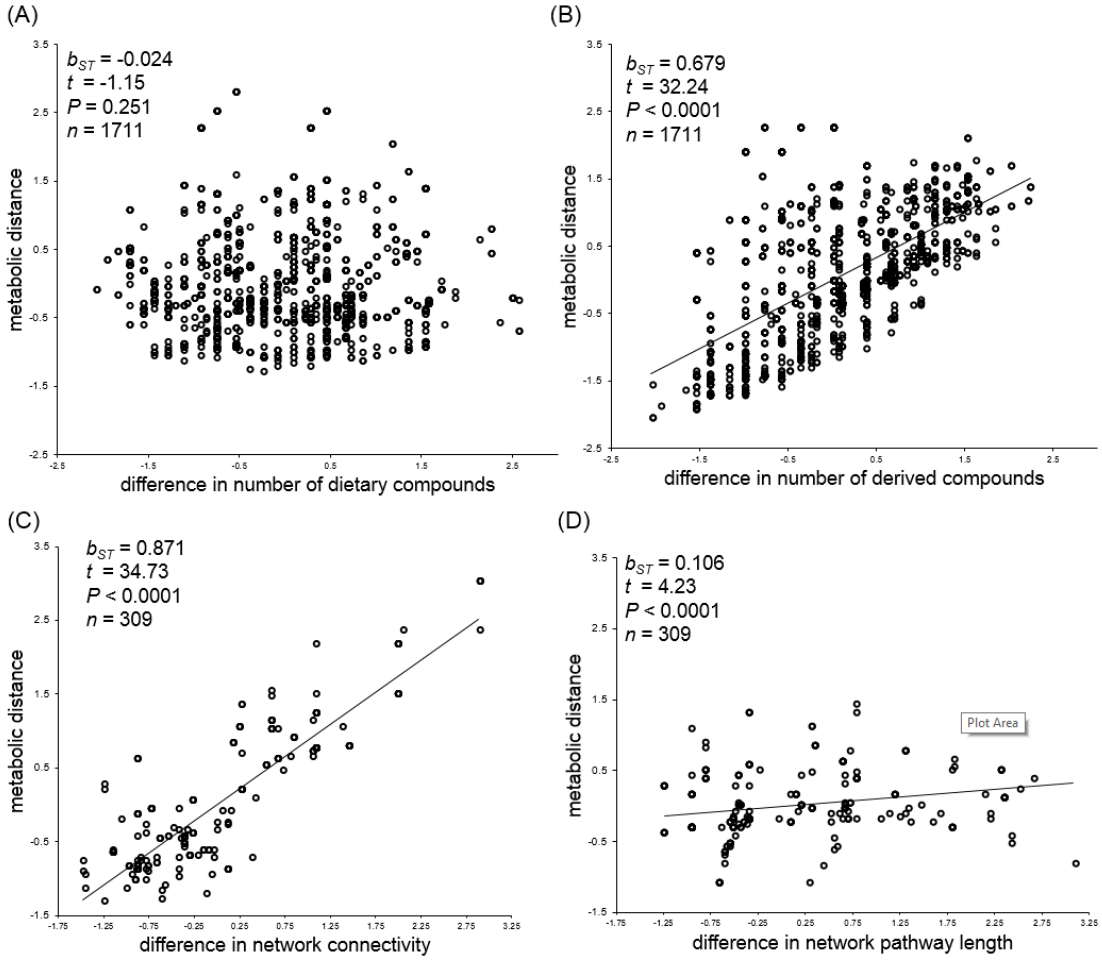

Figure 4: Regressions of pairwise metabolic distance between female networks on (A) the difference in number of dietary compounds and (B) the difference in number of metabolically derived compounds (non-dietary). Regressions of pairwise metabolic distance between female carotenoid networks with the same dietary compounds on (C) the difference in the number of pathways (network connectivity) between each pair of female networks and (D) between the difference in length of pathways between pairs of networks.

REFERENCES

Assenov Y, Ramírez F, Schelhorn S-E, Lengauer T, Albrecht M (2008) Computing topological parameters of biological networks. Bioinformatics 24.2: 282-284.

Badyaev, A.V., Morrison, E.S., Belloni, V., and Sanderson, M.J. (2015) Tradeoff between robustness and elaboration in carotenoid networks produces cycles of avian color diversification. Biology Direct 10: 45.

Badyaev, A.V. and Martin, T.E. (2000) Sexual dimorphism in relation to current selection in the house finch. Evolution 54: 987-997.

Barabasi, Albert-Laszlo, and Zoltan N. Oltvai. (2004) Network Biology: Understanding the Cell's Functional Organization. Nature Reviews Genetics 5: 101-13.

Chi, Jen-Tsan, Zhen Wang, Dimitry S. A Nuyten, Edwin H. Rodriguez, Marci E. Schaner, Ali Salim, Yun Wang, Gunnar B. Kristensen, Åslaug Helland, Anne-Lise Børresen-Dale, Amato Giaccia, Michael T. Longaker, Trevor Hastie, George P. Yang, Marc J Van De Vijver, and Patrick O. Brown. (2006) Gene Expression Programs in Response to Hypoxia: Cell Type Specificity and Prognostic Significance in Human Cancers. PLoS Med PLoS Medicine 3.3: 47.

Díaz-Hernández, Verónica, Alejandro Marmolejo-Valencia, and Horacio Merchant-Larios. (2015) Exogenous Estradiol Alters Gonadal Growth and Timing of Temperature Sex Determination in Gonads of Sea Turtle. Developmental Biology 408.1: 79-89.

Doncheva NT, Assenov Y, Domingues FS, Albrecht M (2012) Topological analysis and interactive visualization of biological networks and protein structures. Nature Protocols 7.4: 670-685.

Greene, E. (1989) A Diet-Induced Developmental Polymorphism in a Caterpillar. Science 243.4891: 643-46.

Higginson, D.H., Belloni, V., Davis, S.N., Morrison, E.S., Andrews, J.E., and Badyaev, A.V. In press. Evolution of long-term coloration tredns with biochemically unstable ingredients. Proc Roy Soc B.

Khan, Raja Amjad Waheed, Jianhua Chen, Meng Wang, Zhiqiang Li, Jiawei Shen, Zujia Wen, Zhijian Song, Wenjin Li, Yifeng Xu, Lishan Wang, and Yongyong Shi. (2016) A New Risk Locus in the ZEB2 Gene for Schizophrenia in the Han Chinese Population. Progress in Neuro-Psychopharmacology and Biological Psychiatry 66: 97-103.

Landeen, E.A. and Badyaev, A.V. (2012) Developmental integration of feather growth and pigmentation and its implications for the evolution of diet-derived coloration. J. Exp. Zool. B Mol. Dev. Evol. 318: 59-70.

Lozier, Jeffrey D., Jason M. Jackson, Michael E. Dillon, and James P. Strange. (2016) Population Genomics of Divergence among Extreme and Intermediate Color Forms in a Polymorphic Insect. Ecol Evol Ecology and Evolution 6.4: 1075-091.

Ma, Hongwu, Anatoly Sorokin, Alexander Mazein, Alex Selkov, Evgeni Selkov, Oleg Demin, and Igor Goryanin. (2007) The Edinburgh Human Metabolic Network Reconstruction and Its Functional Analysis. Mol Syst Biol Molecular Systems Biology 3: 135

McGraw K.(2006) The mechanics of carotenoid coloration in birds. In: Geoffrey E. Hill KJM (ed) Bird Coloration, Volume 1: Mechanisms and Measurements, 1 edn. Harvard University Press, Cambridge, MA. pp. 177-242.

McSweeney PJ (2008) Randomnetworks. Version 1.0. http://apps.cytoscape.org/apps/randomnetworks. In.

Morrison, E.S. and Badyaev, A.V. In press. The landscape of evolution: Reconciling structural and dynamic properties of metabolic networks in adaptive diversifications. Integrative and Comparative Biology.

Morrison, E.S. and Badyaev, A.V. In review. Structuring evolution: Biochemical networks and metabolic diversification in birds.

Segers, Francisca H. I. D., Cristiano Menezes, Ayrton Vollet-Neto, Dorothee Lambert, and Christoph Grüter. (2015) Soldier Production in a Stingless Bee Depends on Rearing Location and Nurse Behaviour. Behavioral Ecology and Sociobiology Behav Ecol Sociobiol 69.4: 613-23.

Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27.3: 431-432.

Storrs, E. E., and R. J. Williams. (1968) Study of Monozygous Quadruplet Armadillos in Relation to Mammalian Inheritance. Science 160.3826: 443.

Sutter, N. B., C. D. Bustamante, K. Chase, M. M. Gray, K. Zhao, L. Zhu, B. Padhukasahasram, E. Karlins, S. Davis, P. G. Jones, P. Quignon, G. S. Johnson, H. G. Parker, N. Fretwell, D. S. Mosher, D. F. Lawler, E. Satyaraj, M. Nordborg, K. G. Lark, R. K. Wayne, and E. A. Ostrander. (2007) A Single IGF1 Allele Is a Major Determinant of Small Size in Dogs. Science 316.5821: 112-15.