Evolutionary Dissection and Regulation of Animal Complex Traits Postprint

Authors: Kunming Institute of Zoology, Chinese Academy of Sciences

Date: 2017-02-08T00:00:00+00:00

Abstract

Animal complex traits represent the outcome of long-term adaptive evolution in animals and constitute the primary foundation for animal diversity. Dysregulation of these traits serves as an intrinsic cause of major chronic diseases in humans, while understanding the principles governing their formation holds significant importance for improving economic traits in domestic animals and for biomimicry of specialized animal functions. Consequently, systematically dissecting the causative mechanisms of animal complex traits constitutes not only a fundamental frontier scientific endeavor aimed at revealing the essential laws of nature, but also has significant application potential. The mechanistic dissection of animal complex trait formation has long remained a scientific challenge, confronted with difficulties in traceability, predictability, and controllability. The pilot program "Evolutionary Dissection and Regulation of Animal Complex Traits" seeks to transcend the current framework of GWAS (Genome-Wide Association Study) through large-scale, cross-species evolutionary comparisons, and to systematically integrate comprehensive data across three dimensions—genetic factors, developmental network evolution, and phenotypic adaptation—in diverse animal species to dissect the causative mechanisms of animal complex traits. We designate this novel approach to studying animal complex traits as eGPS (evolutionary Genotype-Phenotype Systems biology).

Full Text

Preamble

Strategic Priority Research Program (Category B) of the Chinese Academy of Sciences
ChinaXiv Collaborative Journal: Evolutionary Analysis and Regulation of Animal Complex Traits

1. Project Background and Significance

Animal complex traits are the product of long-term adaptive evolution and constitute the primary foundation for biodiversity. Imbalances in their regulation represent the underlying cause of major chronic diseases in humans, while understanding their formation patterns is crucial for improving economic traits in domestic animals and for biomimicry of animal special functions. Therefore, systematically dissecting the origins of animal complex traits is not only fundamental frontier scientific work that reveals the essential laws of nature but also holds significant application prospects.

The analysis of animal complex trait formation mechanisms has long been a scientific challenge, facing difficulties in traceability, predictability, and controllability. The "Evolutionary Analysis and Regulation of Animal Complex Traits" pilot program attempts to address these challenges through large-scale, cross-species evolutionary comparisons that systematically integrate comprehensive data across three dimensions: genetic factors, developmental network evolution, and phenotypic adaptation. By moving beyond the current GWAS (genome-wide association study) framework to dissect the origins of animal complex traits, we have termed this novel research approach eGPS (evolutionary Genotype-Phenotype Systems biology). This approach aims to elucidate the evolutionary patterns of major intrinsic genetic variation forces driving complex trait evolution, establish eGPS models, clarify the causes of animal complex traits, and achieve the goals of making animal complex traits traceable, predictable, and regulable.

2. Progress Achieved

Since its launch in June 2014, the program has achieved a series of important research advances and has completed the original planned tasks. The program has first established large-scale genetic manipulation platforms for non-model animals (monkeys and butterflies) and constructed the world' s most taxonomically comprehensive high-altitude animal genetic resource bank (>20,000 samples with complex trait data). Second, it has overcome conventional method limitations to establish novel methods for complex animal genome sequencing, assembly, and analysis. Additionally, the program has developed methodological systems and models for analyzing multi-level, high-dimensional massive data on animal complex traits.

Building upon these foundations, the program has taken the intrinsic laws underlying the generation of representative complex traits—animal brain size, high-altitude adaptation, and fat energy metabolism—as the main thread running through the entire research effort. From three progressive levels—genetics, developmental regulatory network evolution, and functional adaptation—the program has dissected the origins of animal complex traits. At the genetic level, it has clarified the evolutionary patterns of major intrinsic genetic variation forces driving complex trait evolution. It has then identified the genetic elements determining complex traits such as animal brain size, high-altitude adaptation, and fat energy metabolism. At the developmental network evolution level, it has elucidated how these elements are assembled. At the functional research level, it has primarily used functional experiments to verify results from the previous two levels and understand the adaptive significance of these elements' composition and interactions, as well as the mechanisms leading to maladaptation (such as disease) when equilibrium states are disrupted—namely, analyzing the evolutionary cost or trade-offs of trait evolution.

To date, the program has published 154 papers, including 37 in top-tier international journals such as *Science*, *Nature Genetics*, *PNAS*, *PLoS Biology*, and *Neuron*. Representative achievements are detailed below.

2.1 Establishment of the World' s Most Taxonomically Comprehensive High-Altitude Animal Genetic Resource Bank and Phenotypic Database

To dissect the adaptive evolution mechanisms and energy metabolism regulation of high-altitude animals, the program collected and established the world' s most taxonomically comprehensive high-altitude animal genetic resource sample bank to date. This includes over 3,000 DNA samples from wild vertebrates on the Qinghai-Tibet Plateau (fish, amphibians, reptiles, birds, and mammals) and over 1,000 samples from domestic animals on the plateau (pigs, horses, donkeys, cattle, dogs, and yaks). The program measured and compiled 27 physiological indicators across 66 high-altitude animal populations at different altitudes, including blood cell parameters, blood gas parameters, and heart weight indices. Additionally, 16,206 blood samples from Tibetan populations, 250 placental tissue samples, and 105 primary cultured cell lines from newborn umbilical veins were collected from 65 geographic populations. This work has laid the foundation for constructing eGPS computational and analytical mathematical models and for dissecting the molecular mechanisms of high-altitude adaptation in plateau animals (including humans).

2.2 Major Breakthroughs in the Genetic Mechanisms of Animal High-Altitude Adaptation

Beyond their important conservation biology value, snub-nosed monkey species have developed specialized leaf-eating diets and occupy habitat types ranging from low to high altitudes (800–4,500 m), providing an excellent model for studying the genetic mechanisms of high-altitude environmental adaptation. The program team conducted systematic analyses of the high-altitude environmental adaptation genetics in snub-nosed monkeys, finding significant expansion of genes related to DNA repair and oxidative phosphorylation in *Rhinopithecus bieti*. Among three high-altitude snub-nosed monkey species (*Rhinopithecus bieti*, *Rhinopithecus strykeri*, and *Rhinopithecus roxellana*), eight shared amino acid substitutions were identified in six genes related to lung function, DNA repair, and angiogenesis. The RNASE4 gene variant was confirmed to have higher activity in inducing angiogenesis. This study, using non-human primates as a research model, provides new and more comprehensive insights into high-altitude adaptation as a complex trait.

Using the eGPS research framework, the program team identified several novel genes closely related to high-altitude adaptation and studied their in vivo and in vitro functions. For example, by analyzing genomic and transcriptomic data from plateau antelopes, yaks, zokors, and pikas and comparing them with corresponding plain species data, plateau-specific mutations were discovered in a metabolic enzyme gene of plateau mammals. In vitro cell experiments revealed that the mutant protein had significantly higher enzymatic activity than the wild-type protein. Genetically modified mice carrying this mutation showed significantly longer survival under hypoxic conditions than wild-type mice, and their major organs had higher oxygen-carrying capacity, indicating that this gene may be an important gene related to high-altitude animal adaptation.

The program also systematically dissected the molecular mechanisms of high-altitude adaptation in plateau fish, amphibians and reptiles, and Tibetan chickens. Beyond genomic discoveries, the program identified high-altitude-specific adaptive groups in the microbial flora of plateau yaks and other animals. The eGPS methodology was also applied to brain morphological analysis, revealing that phospholipid remodeling plays a crucial role in human brain function evolution, with lipid enrichment in the human brain being five times higher than in non-nervous system tissues.

2.3 Evolution of Animal Energy Metabolism Mechanisms

Sheep are a major source of meat, milk, and wool fibers and represent a special class of animals that use a specialized digestive organ (the rumen) for preliminary digestion of plant materials. The program collected and obtained high-quality genomic and transcriptomic data from 40 different sheep tissues, discovering that many genes involved in lipid metabolism have undergone expansion or show different tissue expression patterns, likely in response to changes in the skin' s lipid metabolism barrier and wool synthesis [Figure 1: see original paper].

Giant pandas primarily consume bamboo and maintain energy balance by reducing the size of some major organs and decreasing daily activity. Program members measured the energy metabolism of five captive and three wild giant pandas, revealing that a stop codon mutation in the DUOX2 gene coding sequence leads to lower thyroid hormone levels in peripheral blood and lower energy requirements. The program thus provided a comprehensive analysis of the mechanisms enabling giant pandas to survive and maintain energy primarily on a low-energy bamboo diet from morphological, behavioral, physiological, and genetic perspectives [Figure 2: see original paper].

3. Originality and Innovation

(1) Establishment of novel eGPS research paradigms and computational methods. Using cross-species and cross-tissue large-scale data, program team members developed new general quantitative models and computational methods for animal eGPS research (evolutionary genotype-phenotype systems biology). The program developed a new theoretical framework based on Dynamic Network Biomarkers (DNB) to quantify critical transition nodes from one trait to another. It also introduced the concept of Partial Mutual Information (PMI), which can effectively reconstruct large-scale biological information networks such as gene regulatory networks and protein-protein interaction networks from existing omics data. These new methods and models have been applied to dissect the high-altitude adaptation mechanisms in Tibetan populations.

(2) Construction of a non-model animal genetic modification research platform. The program achieved comprehensive breakthroughs in transgenic and gene-editing technologies for monkeys and tree shrews, obtaining the world's largest number of transgenic and gene-knockout monkeys for brain development studies [Figure 3: see original paper]. It successfully established technical systems and experimental platforms for producing transgenic and gene-knockout monkeys using lentiviruses and CRISPR-Cas9 technology, batch-constructing over 40 transgenic or gene-knockout monkeys. Through spermatogonial stem cell replacement, the program achieved the world' s first successful gene editing in tree shrews. Additionally, the program successfully used CRISPR-Cas9 technology to edit genes in butterflies.

(3) Construction of an omics database system. The program established an omics database system with independent Chinese intellectual property rights (Genome Sequence Archive, GSA), which has been recognized by multiple international academic journals. The GSA database system represents a domestic "zero-to-one" breakthrough and will gradually break the monopoly of the International Nucleotide Sequence Database Collaboration (INSDC) on omics data, securing China' s position in international omics data management and sharing. This achievement holds important social value and impact. Furthermore, through the accumulation of data and knowledge, future deep mining and utilization of omics big data will enable extensive applications in biological breeding, population health, major diseases, and other areas, generating significant economic value.

4. Significance for Industry

Diseases can be viewed as a special category of complex traits. The eGPS research system established by this program is valuable for fully mining human disease-related big data, revealing the regulatory networks underlying disease occurrence and development, and providing new diagnostic and therapeutic targets and strategies. The program' s establishment of transgenic genetic manipulation and functional verification platforms for non-model animals (monkeys, tree shrews, etc.) will provide resource guarantees for subsequent functional dissection of brain capacity and structural complexity as complex traits and holds significant importance for creating human disease models and drug development.

5. Recommendations for Future Deployment in Disciplines, Industry Advancement, and Talent Cultivation in China

The rapid development of omics technologies, represented by genomics, has generated massive amounts of data, providing conditions for systematically dissecting the origins of animal complex traits. The eGPS technology established by this program provides an important platform for systematically mining and integrating massive omics data and revealing the regulatory networks of complex traits. From a disciplinary perspective, against the backdrop of the genomics era, the integrated research field of "Genetics, Development, and Evolution" has become one of the international research hotspots, yielding a series of important achievements. In recent years, China' s overall research level in this disciplinary field has improved rapidly and received high attention from international peers. The current period is a critical time for genetics, development, and evolution research and represents the best opportunity for achieving major original breakthroughs. Chinese scientists have made excellent progress and established distinctive features in this area, reaching world-class levels in certain directions.

It is recommended that China concentrate its outstanding research forces in related fields to conduct studies on the molecular mechanisms of biological adaptation at the whole-genome level, focusing on the most significant core scientific questions in the field of biological adaptive evolution and the nation' s most urgent major needs. This will enhance China' s basic innovation capabilities in life sciences and strengthen China' s overall scientific and technological competitiveness.

(Supporting Institution: Kunming Institute of Zoology, Chinese Academy of Sciences)

This is a highly innovative project aimed at dissecting the evolutionary mechanisms of animal complex traits across species. By integrating the most cutting-edge omics analytical approaches through the eGPS methodology, the research team has achieved major breakthroughs in dissecting the evolutionary mechanisms of three representative complex traits: animal brain size, high-altitude adaptation, and fat energy metabolism. The program has published 154 papers, including 37 in top-tier international journals such as *Science*, *Nature Genetics*, *PNAS*, *PLoS Biology*, and *Neuron*. These research findings will provide scientific evidence for dissecting the genetic basis of physiological and morphological traits and will supply basic scientific data for future targeted research on obesity-related human diseases. More importantly, this program will provide an unprecedented methodological platform for dissecting the genetic mechanisms of animal complex traits, including disease occurrence, with particular emphasis on developing models and methods for integrating cross-species, multi-omics data, which is of great significance. If further integrated and developed into a complete theoretical system and methodological framework, it will also have important reference value for research in other areas of life sciences.

Li Peng, Academician of the Chinese Academy of Sciences, Academician of the World Academy of Sciences for the advancement of science in developing countries, Professor at the School of Life Sciences, Tsinghua University, Director of the Life Sciences Division of the National Natural Science Foundation of China, Chang Jiang Scholar Distinguished Professor, Leader of the Chang Jiang Scholar Innovation Team, Expert receiving Special Government Allowances of the State Council, Chief Scientist of the "973" Program, and recipient of the National Science Fund for Distinguished Young Scholars. Her primary research focuses on the regulatory mechanisms of lipid metabolism and the pathogenesis of metabolic diseases such as obesity, diabetes, and fatty liver disease. She has published over influential articles in top international journals including *Cell*, *Nature Medicine*, *Nature Cell Biology*, *Cell Metabolism*, and *Genes & Development*.

Through the hard work of the first two years, the program has achieved outstanding results overall, including publishing high-level academic papers and establishing original theories, methods, and technologies, such as the novel eGPS research paradigm and computational methods, the non-model animal genetic modification research platform, and the omics database system (Genome Sequence Archive, GSA). The program has made pioneering discoveries using the eGPS method in non-model animals regarding energy balance and fat storage regulation mechanisms, with three important papers published in *Science* and one in *PLoS Genetics*. Major findings include evidence regarding cheetah predatory activity and energy expenditure, the adaptive mechanisms enabling giant pandas to survive on low-energy bamboo diets, novel lipid metabolism pathways in sheep skin, and new genetic studies of lipid signaling in insects and other arthropods.

Yang Hongyuan, Professor at the University of New South Wales, Australia. His primary research focuses on the mechanisms underlying the development and progression of metabolic diseases, including adipose development, the molecular mechanisms of fat accumulation, secretion, and breakdown in adipocytes and liver cells, and their relationship with insulin sensitivity, as well as intracellular lipid transport. In recent years, he has published 68 articles in various journals including *Cell*, *Cell Reports*, *Diabetes*, and *Journal of Cell Biology*.

Note: Figure translations are in progress. See original paper for figures.