Summary

Yoshinori Endo

Background

Biodiversity is rapidly declining worldwide. Understanding of biological functions of wild animals have revealed evolution of species unique biological functions and enabled effective conservation management. Although a whole genome sequencing has become common among non-model species, it is challenging to understand the evolution of biological function at large-scale study. Although cellular experiments enable to investigate the functional influence of the genetic changes to phenotypic characteristics, viable cellular material is often limited in wild animals because of ethical and technical concern, the advent of induced pluripotent stem cell (iPSCs), that can grow indefinitely and differentiate into any types of cells, has now made possible to obtain variety of and infinite cellular resources. However, iPSCs generated from range of taxonomic groups have exhibited variety of characteristics, suggesting development of unique pluripotency mechanisms. Here, I accomplished comparative evolutionary approach to gain insights into genetic changes underlying mammalian biological functions, especially focusing on pluripotent stem cells.

Genetic signatures underlying development of species' unique characteristics is essential for understanding of evolution. In mammalian evolutionary history, Cetacea (whales, dolphins, and porpoises) achieved astonishing success by adapting to an aquatic environment, making them a poster child for evolutionary study. One unique characteristic of cetaceans, contributing to this

adaptive success, is efficient lipid utilization. To infer the molecular changes underlying their lipid utilization, I investigated lipid metabolism associated genes in Cetacea and Bovidae.

Genetic changes may also enable insights into variations of mammalian PSCs. PSCs from different species show variations in some of their characteristics such as pluripotent state and gene expression. Various studies have shown molecular mechanisms underlying the maintenance of PSC characteristics that constitute a complex gene regulatory network. Unfortunately, such knowledge is largely limited to primates and rodents. To this end, I performed comparative analyses involving pluripotency-regulating genes across mammalian taxa. As such, it helps the understanding of the genetic basis for the evolution of the mammalian PSCs.

Variations of mammalian PSCs were also investigated through generation of iPSCs from new species. Generation of iPSCs from new species will enable insights into species differences and similarities in characteristic of mammalian PSCs. From a perspective of biodiversity conservation, iPSC technology has opened new avenue for saving the endangered species. Once established, iPSCs from rare species are potentially useful for therapeutic applications and reproduction of individuals through germline differentiation. Herein, I report generation and characterization of iPSCs from new endangered species, the Grevy's zebra.

Methods

The genetic signatures in evolution of characteristics through adaptation was investigated by a comparative genetic analysis of five aquatic and five terrestrial Cetartiodactyla species. 144 lipid metabolism associated genes were collected for each species. Mutation ratio (d_N/d_S), amino acid substitution in functional domains, and metabolic pathways were evaluated using branch-site model in PAML, Pfam, and KEGG, respectively. Evolutionary inferences were explored here for the relationships between positively selected genes and adaptive biological functions of Cetacea.

The evolution of pluripotency regulating network was investigated through genetic conservation and variations in pluripotency related genes across mammals. Comparative genetic analysis was performed with 134 genes constituting the pluripotency gene regulatory network across 48 mammalian species covering all the major taxonomic groups. The stringency of negative selection on pluripotency genes were tested by RELAX. Evolutionary rate of each gene was estimated based on the dN/dS ratio, then the conservation patterns of the network subcircuits were compared with ANOVA, followed by the Tukey-Kramer. For the genes involved in the subcircuit with high evolutionary rate, lineage-specific positive selection was tested using Branch-site model in PAML. The positively selected genes were inferred for variations of PSC characteristics and species unique biological functions.

Grevy's zebra skin tissue was collected from a dead individual a zoo. The reprogramming of the fibroblasts was performed by transducing four transcription factors, *Oct3/4*, *Sox2*, *Klf4*, and

c-Myc, using retroviral vectors. Zebra iPSCs were culture condition developed for human iPSCs.

Pluripotency of zebra iPSCs were tested according to standard criteria developed in human and

mouse. The characteristics of zebra iPSCs were observed in their morphology, culture condition,

and gene expressions. RNA-seq was used for transcriptome analysis.

Results

The study of lipid metabolism evolution revealed unique and divergent evolution of this biological

function in Cetacea and Bovidae. Positive selection analysis detected 20 positively selected genes

in Cetacea compared to 11 in Bovidae with little overlap between the lineages. I identified lineage-

specific patterns of amino acid substitutions and functional domains that were mutually exclusive

between cetaceans and bovids. Moreover, a pathway analysis showed that the identified genes in

cetaceans were associated with lipid digestion, lipid storage, and energy producing pathways.

Mammalian genes in the pluripotency regulatory network show a remarkably high degree

of evolutionary stasis, especially in the core regulatory elements of the network, Jak-STAT

signaling, and PI3K-Akt signaling pathways. Nevertheless, despite the overall conservation of the

regulatory network, the downstream targets of the core regulatory elements showed relatively high

evolutionary rate and contained multiples genes that have undergone positive selection. Those

positively selected genes include cancer related genes in large bodies animals and placenta related

4

genes in Eutheria. I further revealed that positively selected genes could be associated with species'

unique adaptive characteristics.

Colonies with iPSC-like morphology appeared by transducing zebra fibroblasts using

human transcription factors. Zebra iPSCs in this study showed primed-type morphology and can

be maintained with primed-type culture condition, as human iPSCs. Zebra iPSCs expressed

pluripotency markers at genetic and protein levels. Transcriptome of zebra iPSCs were nested

within PSCs from other mammalian species.

Discussion

In the study of lipid metabolism evolution in Cetacea, the little overlap of positively selected genes

between Cetacea and Bovidae indicates independent and magnified changes in Cetacean lipid

metabolism since divergence from the terrestrial common ancestor between Bovidae. The lineage-

specific patterns of amino acid substitutions and functional domains suggests these molecular

changes in Cetacea are corresponding to their aquatic adaptation. The lipid metabolism pathways

associated with the positively selected genes were consistent with adaptive characteristics under

aquatic environment, implying development of unique characteristics in these functions of Cetacea.

This study emphasizes the evolutionary context of lipid metabolism modification of cetaceans and

provides a foundation for future studies of elucidating the adapted biological mechanisms of

cetacean lipid metabolism and a framework for incorporating ecological context into studies aimed at investigating adaptive evolution.

The study of pluripotency gene regulating network provided new insights into evolutionary conservation and variation within pluripotency genes and their network. The high degree of genetic conservation among pathways which participates in proliferation and self-renewability suggests the conservation of fundamental biological process of mammalian PSCs across species. On the other hand, positive selections with multiple pluripotency genes supports the development of species variations in mammalian pluripotent stem cells. The associations between the positively selected genes and species unique biology indicates that species' adaptation might be a possible driver for variations of mammalian PSCs. These results provide important insight into the evolution of the pluripotency gene regulatory network underlying variations in characteristics of mammalian PSCs.

The expression of markers for pluripotency and transcriptome indicates the zebra fibroblasts were successfully reprogrammed into iPSCs. The generation of zebra iPSCs with human transcription factors suggests conservation of reprogramming mechanism between these taxonomically distant species. Because viable cellular materials enable to investigate biological characteristics rather than protein-coding genes, comparative transcriptomic analysis with other mammalian species provided further insights into variation of mammalian PSCs. Zebra iPSCs

established in this study will provide a foundation for future application of the iPSC technology to conservation management.

The comparative evolutionary study provided new insights into genetic changes underlying mammalian biological functions. By focusing genomic scale, I revealed the lineage and cellular specific genetic changes that were not described previously. Identification of genetic signatures of particular biological function is essential for functional study, especially when the detected genetic changes will be further investigated with cellular experiment. The new insights of variation of pluripotent stem cells found in this study will facilitate the application of iPSC technology to wide range of mammalian species, including endangered species. The application of iPSC technology to range of species will contribute to elucidate the principle of pluripotent stem cells. Furthermore, functional studies of wild animals using iPSC-derived differentiated cells will shed light on understanding of evolution of our own biology.