Lengthening of 3'UTR Increases Morphological Complexity in Animal Evolution

Cho-Yi Chen[1,2,3], Shui-Tein Chen[2], Hsueh-Fen Juan[3*], Hsuan-Cheng Huang[1*]

[1]Institute of Biomedical Informatics, Center for Systems and Synthetic Biology, National Yang-Ming University, Taipei, Taiwan, [2]Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan, [3]Genome and Systems Biology Degree Program, Department of Life Science, Institute of Molecular and Cellular Biology, Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan.

[*]To whom correspondence should be addressed. E-mail: yukijuan@ntu.edu.tw, hsuancheng@ym.edu.tw.

Abstract

By analyzing the structure of mRNA transcripts in multiple metazoan species, we observed a striking exponential correlation between the length of 3' untranslated regions (3'UTR) and morphological complexity as measured by the number of cell types in each organism. Cellular diversity was similarly associated with the accumulation of microRNA genes and their putative targets. We propose that the lengthening of 3'UTRs together with a commensurate expansion in post-transcriptional regulation can contribute to the emergence of new cell types during animal evolution.

Main Text

Over the last decade of genome exploration, research into the evolution of organismal diversity and morphological complexity has shifted its focus from raw genome size to gene regulatory complexity[1-3]. Complex gene regulatory programs may allow a greater variety of potential combinations of gene expression, thereby facilitating the generation of novel cell types[1]. Here, we investigate whether this principle also applies at the post-transcriptional level.

After analyzing the structure of genomic mRNA transcripts from a variety of multicellular animals (see **Supplementary Methods** online), we observed a striking exponential correlation (Pearson's $\rho = 0.92$, $P = 1.39e\text{-}04$) between the typical length of the 3' untranslated region (3'UTR) and morphological complexity as measured by the number of cell types (**Fig. 1a**). Conversely, there was no significant correlation observed between morphological complexity and 5'UTR length (**Supplementary Fig. 1a** online) or the length of the coding sequence (**Supplementary Fig. 1b** online). Since animal miRNAs mainly target the *cis*-regulatory elements in 3'UTR, this difference may be due to miRNA-based post-transcriptional regulation[4]. Indeed, miRNAs play a crucial role in the evolution of 3'UTR and the establishment of tissue identity[5-8]. The likelihood that a transcript will be targeted by miRNA may increase with 3'UTR length because 3'UTR expansion can yield additional miRNA binding sites[9]. Furthermore, longer 3'UTRs may include multiple polyadenylation signals and thus produce several transcript isoforms with different combinations of miRNA binding sites, leading to a large increase in regulatory complexity[9].

The number of miRNA genes in the genomes was also found to grow exponentially with morphological complexity (Pearson's $\rho = 0.77$, $P = 9.26e\text{-}03$; **Fig. 1b**). The co-expansion of *trans*-regulators (miRNA genes) and potential *cis*-elements (length of 3'UTR) suggests a possible evolutionary increase in the complexity of post-transcriptional regulation. Indeed, we found that the miRNA binding complexity in 3'UTR, as predicted by different algorithms (**Supplementary Methods** online), strongly correlated with morphological complexity (**Fig. 1c**).

Lengthening of 3'UTR Increases Morphological Complexity

In addition, the typical miRNA target and anti-target categories have a clear divergence in 3'UTR length patterns (**Fig. 1d**). The length of 3'UTRs in transcriptional regulation genes, which represent a typical class of miRNA targets[6], grew exponentially as organismal complexity increased. In sharp contrast to this, ribosomal genes, a class of housekeeping genes and miRNA anti-targets[6], had relatively short 3'UTRs in all model organisms. These results reinforce the argument that the lengthening of 3'UTR is highly associated with the expansion of miRNA regulatory programs.

It should be noted that a similar exponential correlation of morphological complexity was demonstrated in the number of transcription factors (TF) (Pearson's $\rho = 0.70$, $P = 1.81e\text{-}03$) (**Supplementary Fig. 2b** online). This suggests that the expansion of gene regulators occurred and coevolved at both the transcriptional and post-transcriptional levels. Given the role of miRNAs in stabilizing gene expression and conferring robustness to gene regulatory networks[6,8], they are likely to play an important role in establishing stable expression landscapes and maintaining tissue identities in cooperation with TFs.

Interestingly, all the regulation-related variables (L) examined above showed similar exponential growth against the number of cell types (n): $L(n) \sim e^{\alpha n}$, or equivalently, $dL/dn \sim \alpha L$, with a factor α ranging from 0.007 to 0.016. This implies that the generation of a new cell type involves an approximate 1% increase in regulatory complexity.

These findings align with recent studies showing that protein domain recombination and *cis*-regulatory element recombination may have a major role in evolution[10-11]. These changes occur in the 3'UTR and can open up new miRNA binding motifs[9]. Since gene duplication and recombination are combinatorial processes that both rely on existing genetic materials, it follows that 'complexity will breed complexity', leading to a non-linear expansion in post-transcriptional interactions. Based on our empirical evidence of a strong, exponential trend between the development of new cell types and the number of gene regulators, we therefore propose that 3'UTR lengthening can lead to greater morphological complexity in animals and that this process will be driven by natural selection.

Acknowledgements

Lengthening of 3'UTR Increases Morphological Complexity

We thank B.-Y. Liao and F.-C. Chen for helpful discussion. This work was supported by the National Science Council (NSC 99-2621-B- 010-001-MY3 and 002-005-MY3) and National Health Research Institute (NHRI-EX98-9819PI).

Author Contributions

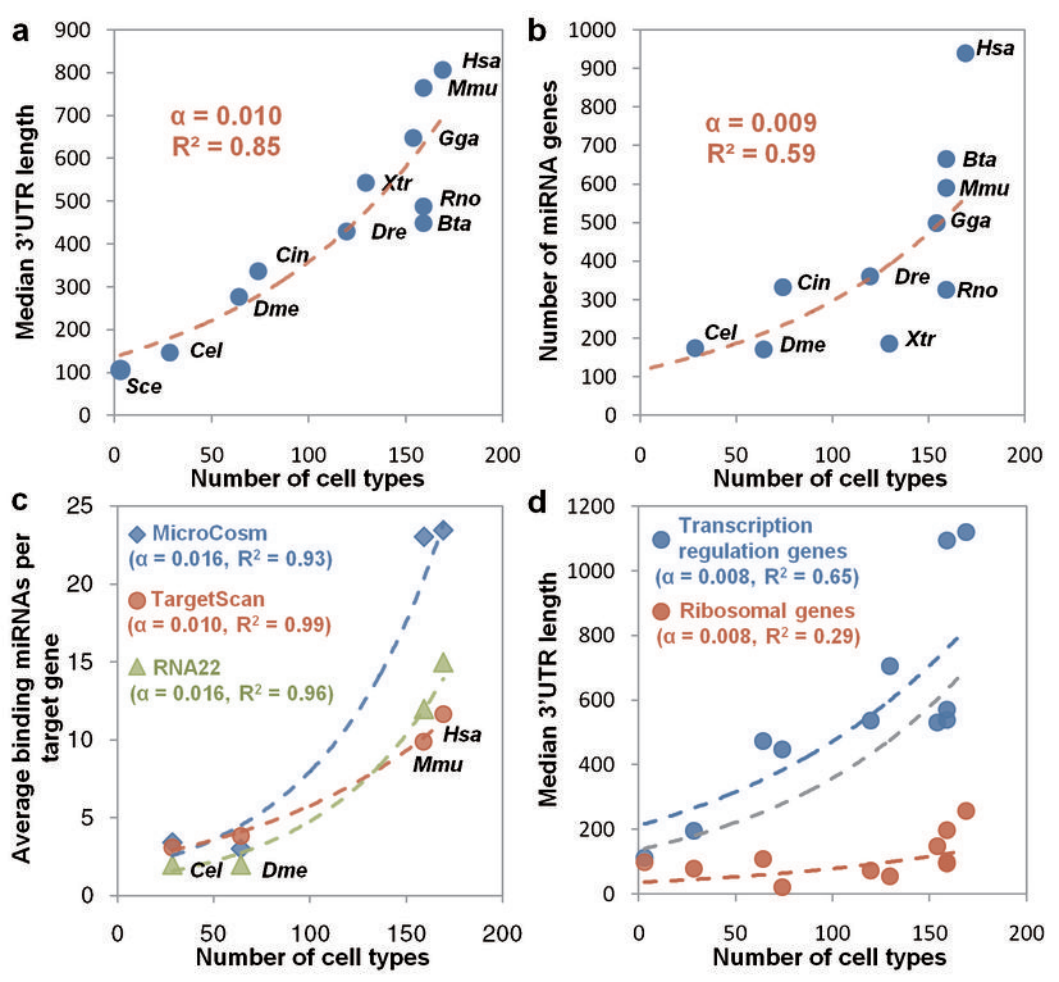
C.Y.C., S.T.C., J.H.F., and H.C.H. designed research and wrote the paper. C.Y.C. and H.C.H. analzed the data.

References

1. Carroll, S.B. Nature 409, 1102-9 (2001).
2. Vogel, C. & Chothia, C. PLoS Comput Biol 2, e48 (2006).
3. Tan, C.S.H. et al. Science 325, 1686-1688 (2009).
4. Mazumder, B., Seshadri, V. & Fox, P.L. Trends Biochem Sci 28, 91-8 (2003).
5. Farh, K.K. et al. Science 310, 1817-21 (2005).
6. Stark, A., Brennecke, J., Bushati, N., Russell, R.B. & Cohen, S.M. Cell 123, 1133-46 (2005).
7. Kosik, K.S. Nat Rev Neurosci 10, 754-759 (2009).
8. De Mulder, K. & Berezikov, E. Genome Biology 11, 111 (2010).
9. Ji, Z., Lee, J.Y., Pan, Z., Jiang, B. & Tian, B. Proc Natl Acad Sci U S A 106, 7028-33 (2009).
10. Peisajovich, S.G., Garbarino, J.E., Wei, P. & Lim, W.A. Science 328, 368-372 (2010).
11. Paixão, T. & Azevedo, R.B.R. PLoS Comput Biol 6, e1000848 (2010).

Figure Legends

Figure 1 Exponential correlation between miRNA-mediated regulation and morphological complexity. (**a**) A strong exponential correlation exists between the median length of 3'UTRs and morphological complexity, as measured by distinct cell types[2], among ten animal species. Budding yeast (*S. cerevisiae*) is included for comparison as a unicellular eukaryote. The dashed line indicates a single exponential fit. (**b**) The number of miRNA genes in each genome and (**c**) miRNA binding complexity (average numbers of putative binding miRNAs per target gene) correlate exponentially with morphological complexity. (**d**) The growth profiles of median 3'UTR length for transcription regulation genes (blue line) and ribosomal genes (red line) have different trends. Budding yeast is included for comparison. The gray line, showing the global trend, is adapted from (**a**).


a
Median 3'UTR length
Number of cell types
α = 0.010
R² = 0.85
Hsa
Mmu
Gga
Xtr
Rno
Bta
Dre
Cin
Dme
Cel
Sce
b
Number of miRNA genes
Number of cell types
α = 0.009
R² = 0.59
Hsa
Bta
Mmu
Gga
Cin
Dre
Rno
Cel
Dme
Xtr
c
Average binding miRNAs per target gene
Number of cell types
MicroCosm
(α = 0.016, R² = 0.93)
TargetScan
(α = 0.010, R² = 0.99)
RNA22
(α = 0.016, R² = 0.96)
Hsa
Mmu
Cel
Dme
d
Median 3'UTR length
Number of cell types
Transcription regulation genes
(α = 0.008, R² = 0.65)
Ribosomal genes
(α = 0.008, R² = 0.29)