

Research Abstract

Submitted May 18, 2012

Title: Comparative genome-wide analysis of transcription initiation and promoter architecture in eukaryotes

Authors: R. Taylor Raborn¹, Krishnakumar Sridharan², Daniel S. Standage², John M. Logsdon Jr.¹, Volker Brendel³

Affiliations:

¹Department of Biology and Roy J. Carver Center for Comparative Genomics, University of Iowa, Iowa City, IA USA 52242

²Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA USA 50011

³Departments of Biology and Informatics, Indiana University, Bloomington, Indiana 47405

Abstract Text:

Accurate annotation and prediction of *cis*-regulatory sequences within genomes is a major goal of computational biology. Given the current abundance of transcriptome sequence data, recent studies have considered positions of transcription initiation, namely transcription start sites (TSSs), to annotate and ultimately characterize promoters for a given genome. However, characterizations of promoter architecture with TSS have been heretofore restricted to metazoans. As such, the current sample is inadequate to comprehensively address the evolution of eukaryotic promoter architecture.

Seeking to address this gap in knowledge, we embarked upon a comparative analysis of promoter architecture using a diverse array of 13 species in four kingdoms of eukaryotes (Table 1), utilizing TSS data deriving from either CAGE or EST datasets. This work has three major objectives: 1) To computationally define the position, breadth and shape of promoters for genes within each species (see Figure 1), 2) To directly compare the annotated promoter architectures between species for orthologous genes, and finally 3) To generate a more complete picture of core promoter composition, potentially providing a hypothesis concerning its evolution across eukaryotes.

To accomplish these objectives, we apply a novel computational tool (TSRchitect), developed to identify and annotate Transcription Start Regions (TSRs) from global TSS studies. Using TSRchitect, we report efficient and accurate definition of TSRs using transcriptome data. Once complete, this work aims to provide a comprehensive analysis of core promoter architecture across eukaryotic diversity.

Appendix

Table 1: List of taxa analyzed in this study.

Common Name	Scientific Name	Kingdom
Human	<i>Homo sapiens</i>	Animalia
House Mouse	<i>Mus musculus</i>	Animalia
Zebrafish	<i>Danio rerio</i>	Animalia
Fruit Fly	<i>Drosophila melanogaster</i>	Animalia
Budding Yeast	<i>Saccharomyces cerevisiae</i>	Fungi
Gray Shag Mushroom	<i>Coprinopsis cinerea</i>	Fungi
Shiitake Mushroom	<i>Lentinula edodes</i>	Fungi
Red Algae	<i>Cyanidioschyzon merolae</i>	Plantae
Thale Cross	<i>Arabidopsis thaliana</i>	Plantae
Soybean	<i>Glycine max</i>	Plantae
Rice	<i>Oryza Sativasativa</i>	Plantae
Toxoplasma	<i>Toxoplasma gondii</i>	Protista
Plasmodium	<i>Plasmodium falciparum</i>	Protista

