Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Comparative studies of gene regulatory mechanisms

Athma A Pai[1] and Yoav Gilad[2]

It has become increasingly clear that changes in gene regulation have played an important role in adaptive evolution both between and within species. Over the past five years, comparative studies have moved beyond simple characterizations of differences in gene expression levels within and between species to studying variation in regulatory mechanisms. We still know relatively little about the precise chain of events that lead to most regulatory adaptations, but we have taken significant steps towards understanding the relative importance of changes in different mechanisms of gene regulatory evolution. In this review, we first discuss insights from comparative studies in model organisms, where the available experimental toolkit is extensive. We then focus on a few recent comparative studies in primates, where the limited feasibility of experimental manipulation dictates the approaches that can be used to study gene regulatory evolution.

Addresses
[1] Department of Biology, Massachusetts Institute of Technology, United States
[2] Department of Human Genetics, University of Chicago, United States

Corresponding author: Gilad, Yoav (gilad@uchicago.edu)

Current Opinion in Genetics & Development 2014, **29**:68–74

This review comes from a themed issue on **Genetics of human origin**

Edited by **Aida Andes** and **Katja Nowick**

http://dx.doi.org/10.1016/j.gde.2014.08.010

Introduction

The controversy over whether changes in gene regulation are disproportionally important in speciation and adaptation relative to changes in protein coding sequences has not yet been resolved [1–3]. Regardless, it has become clear that across a wide range of species, a large number of adaptations can be explained by changes in gene expression levels [4–9]. Similarly, the related question of whether most inter-species differences in gene expression levels have evolved neutrally or were subjected to selective pressures is still unanswered [10]. However, comparative and functional studies of gene expression levels have resulted in a better appreciation of the patterns of regulatory variation within and between species [11–13], and it is now possible to point to subsets of genes whose expression have likely evolved under lineage-specific directional selection [11,12]. The next natural step is to focus on characterizing the underlying regulatory mechanisms.

Broadly speaking, differences in gene expression levels are due to changes in *cis* and/or *trans* regulatory mechanisms [14]. Regulatory elements that act in *cis* (namely, elements that influence allele-specific regulation) include binding sites for transcription factors and small RNAs, as well as sites for chromatin modifiers and marks that determine nucleosome positioning or the degree of chromatin accessibility. Regulatory elements that act in *trans* (namely, affect the regulation of both alleles) include transcription factors and small RNAs, as well as enzymes that modify chromatin and establish epigenetic marks.

Both *cis* and *trans* elements can regulate steady-state gene expression levels by affecting the rates of either transcription or RNA decay. Yet, it has been shown that variation in transcription rates likely accounts for the majority of the overall variation in steady-state transcript levels [15]. Moreover, it has been argued that changes in *cis* might underlie phenotypic adaptations more often than changes in *trans*, since changes to *cis* regulatory elements could be restricted to specific spatial and temporal consequences while changes in *trans* are likely to be associated with general pleiotropic, and often deleterious, effects [14].

Consistent with this notion, we know of a few dozen cases of adaptations in different species that could be explained by changes in gene expression due to genetic variation in *cis* regulatory elements (e.g. [4–9]). Yet we know of a very small number of cases of species-specific regulatory adaptations that can be explained by changes in *trans* elements [16]. In humans, for example, one of the best-characterized cases of possible regulatory adaptation through a *cis* element involves the human-accelerated non-coding sequence 1 (HACNS1), an enhancer region in which human-specific fixed substitutions were shown to drive limb bud expression of nearby genes with possible consequences for human limb development [17•]. In contrast, there are no convincing reports yet of human-specific *trans* regulatory adaptations (though one could arguably consider the accelerated evolution of the human *FOXP2* gene as a possible example [18]). This discrepancy might be partly explained by the inherent difficulty of studying the consequences of suspected adaptive changes in *trans* elements.

Comparative studies of *cis* and *trans* elements

In model organisms and species in which experimentation is feasible, the focus of comparative studies is

typically to uncover the genetic and gene regulatory basis for phenotypic adaptations (these studies are not reviewed here). However, a few studies took advantage of the ability to design specific experiments in model organisms to directly address the question of the relative importance of changes in *cis* and *trans* regulatory mechanisms to the evolution of gene expression. The commonly used approach is to compare RNA sequencing based estimates of allele-specific expression (ASE) levels in F1 hybrids to overall gene expression levels in the homozygote F0 parents.

Using this study design (Figure 1), it is possible to infer whether gene expression differences between the parents are due to *cis*-acting genetic differences that affect allele-specific expression or due to genetic differences that affect both alleles in the F1s, namely differences in *trans* elements. Though the approach does not allow one to easily identify the specific causal regulatory sequence elements, studies using this paradigm took some of the first steps towards deciphering the logic of gene regulatory evolution. Such studies in both flies and mice have suggested that most gene expression differences between strains or closely related species are due to changes in *cis* regulatory elements [19•,20]. These studies have also uncovered substantial and previously under-appreciated contributions from changes in *trans* elements (often acting in combination with changes in *cis*) to inter-strain and inter-species regulatory variation [21,22,23].

In contrast to model organisms, comparative studies in humans and other primates have used indirect approaches to study the relative importance of changes in different regulatory mechanisms, since direct experimentation or hybrid approaches are impossible for ethical and practical reasons. Indeed, changes in *cis* and *trans* regulatory elements in primates were generally inferred based on comparative chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) data. A few recent studies used ChIP-seq data to compare the binding profiles of individual transcription factors (TFs) to identify inter-species differences in binding, and binding site turnover, for specific factors [24–27]. These studies have generally revealed little conservation of TF binding profiles across even closely related primate species [24,25,26,28]. Turnover of TF binding sites is frequent, and even when a TF is bound to orthologous promoter regions across species, it is often not bound to the exactly orthologous regulatory locus [24,25,26,28].

A potential caveat is that nearly all of the comparative ChIP-seq studies published to date have focused on TFs with broad functions, whose binding patterns are extensive (tens of thousands of binding sites genome wide; [25–27]). It is likely that most of these binding events are not directly functional (in the sense that they do not lead to changes in gene regulation [29]) and hence are not expected to be conserved.

An alternative approach is to use comparative profiling of chromatin accessibility, which measures broader differences in putative regulatory elements. This can be done using techniques such as the DNaseI hypersensitivity assay (DNase-seq), which is a genome-wide chromatin accessibility assay that involves the digestion of DNA in regions of open chromatin by the DNaseI enzyme followed by high-throughput sequencing of resulting fragments of accessible regions. DNaseI cleavage sites in open chromatin mark regions that are likely to be regulatory active [30]. Moreover, chromatin accessibility can be used to simultaneously infer the binding of many transcription factors, using DNaseI footprints within accessible regions, using a single assay per individual [31]. The extent of chromatin accessibility as assayed by DNase-seq is generally correlated with gene expression differences across genes and individuals within a species [32–34]. In a comparative context, Shibata *et al.* showed that DNaseI sensitivity differences might explain a modest proportion of differential expression across primates [35].

It should be noted that incomplete power to detect binding events, in either the comparative ChIP-seq or DNase-seq studies, can result in an inflation of apparent inter-species differences (or lack of conservation).

Explaining inter-species variation in gene expression levels

With the advent of better and cheaper high-throughput sequencing technologies, it became possible to characterize genome-wide variation within and between species in a large number of genetic and epigenetic regulatory mechanisms. Although the ultimate goal remains to be able to 'read the code', namely to figure out how the sequences at regulatory elements determine gene expression patterns, an intermediate aim of studies of regulatory evolution is to identify the changes in specific regulatory mechanisms that explain inter-species variation in gene expression levels.

Unfortunately, comparative genome-wide studies are limited in their ability to infer direct causality because they rely on correlations between datasets. Indeed, the general approach has been to characterize inter-species differences in gene expression levels using RNA sequencing, along with variation in one or more regulatory mechanisms. The assumption, based on the central dogma, is that gene expression levels are the output, and changes in regulatory interactions and/or mechanisms are the cause for differences in the output levels. Correlations between inter-species variation in gene expression levels and differences in regulatory mechanisms between species are therefore interpreted as likely to indicate causality. Almost all comparative studies in