Gene Expression Analysis

David S. Skibbe and Virginia Walbot

Abstract A brief history of methods used to elucidate protein function, protein presence, and RNA transcript presence is provided. Gene expression profiling through microarray hybridization, high throughput sequencing, or quantitative reverse transcriptase-polymerase chain reaction methods are reviewed and compared. Proteomics analysis using two dimensional gel electrophoresis followed by protein identification by mass spectrometry is then discussed. Relative costs and prospects for future improvements are also presented.

1 Introduction

To understand how the genome programs plant physiology and development, it is crucial to determine where and when each gene is expressed at the RNA and protein levels and to determine the function(s) of the encoded protein products. The purpose of this Chapter is to review the current methods of gene expression analysis available for maize and to consider methods now in development that will be applicable in the near future.

2 Biochemical Functions of Proteins

Historically, protein functions were more accessible than information about RNA and protein distribution. Many biochemical pathways were worked out before the discovery of messenger RNA! Protein function analysis traditionally required

D.S. Skibbe
Stanford University, Department of Biology
skibbe@stanford.edu

Virginia Walbot
Stanford University, Department of Biology
walbot@stanford.edu
(partial) protein purification, a feat requiring strong skills in biochemistry, and developing a suitable assay. For these reasons, most proteins with well-defined functions are enzymes with readily assayed activities or structural constituents of readily purified complexes such as ribosomes. Today putative functional assignments are made based on sequence similarity to proteins of known function in other organisms, however, the “gold standard” remains a biochemical demonstration of function. The assignment of dehydrogenase, for example, may be correct, but is insufficient without knowing the substrate(s) metabolized by the enzyme. Therefore, it is important to remember that in most cases sophisticated transcriptome and proteome analyses do not resolve protein functions. Of course, maize genetics has the advantage that many loci are studied because they confer a recognizable phenotype; careful phenotypic analysis can refine where and when the gene product is required and genetic analysis can uncover epistatic interactions with other genes and ultimately connect an unknown gene product to known pathways.

Although we will not discuss protein function analysis in any detail, it is clearly an area in which new breakthroughs of protein purification, protein production with appropriate post-translational modifications in transgenic hosts, availability of antibodies to maize proteins, and scaled-up methods for determining functions are required for there to be rapid progress. Today it is much more facile to quantify RNA abundances and distribution (the transcriptome), followed by the amount and distribution of proteins and post-translationally modified isoforms (the proteome), while identification of precise protein functions remains a laborious procedure. High throughput, massively parallel methods for assessing the transcriptome and proteome have made information about gene expression activity and cellular protein constituents relatively inexpensive and accessible to those with only modest expertise in biochemistry.

3 Transcriptome Profiling

There are multiple steps in the RNA-related components of gene expression: primary transcript synthesis dependent on the rate of transcription, processing (such as 5′ capping, intron splicing, polyadenylation at the 3′ end, and nuclear export), translation (reflecting the proportion of cytoplasmic transcript engaged with ribosomes), and finally mRNA turnover. The easiest aspect of gene expression to quantify is the abundance of poly(A)+ mRNA; these transcripts are primarily found in the cytoplasm and most are assumed to be available for or being actively translated. Nonetheless, it is important to remember that quantifying mRNA abundance does not define transcription rate and may or may not be correlated with the abundance of the encoded protein. Perdurance, the persistence of mRNA beyond the time at which it will be translated, will cause a gene to be considered expressed (and by implication important) at stages after its true time of action.

Maize is a very favorable species for transcriptome profiling because of the large size of the plant body and of the individual organs. It is relatively quick to dissect