Journal of Biotechnology 232 (2016) 69–78

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Review

The complexity of bacterial transcriptomes

D. Stazic, B. Voß*

University of Freiburg, Faculty of Biology, Computational Transcriptomics, Schänzlestr. 1, 79104 Freiburg, Germany

ARTICLE INFO

Article history:
Received 29 May 2015
Received in revised form 7 September 2015
Accepted 29 September 2015
Available online 9 October 2015

Keywords:
Transcriptome
Non-coding RNA
Antisense RNA
RNAse
RNA degradation

ABSTRACT

For eukaryotes there seems to be no doubt that differences on the trancriptomic level substantially contribute to the process of species diversification, whereas for bacteria this is thought to be less important. Recent years saw a significant increase in full transcriptome studies for bacteria, which provided deep insight into the architecture of bacterial transcriptomes. Most notably, it became evident that, in contrast to previous scientific consensus, bacterial transcriptomes are quite complex. There exist a large number of *cis*-antisense RNAs, non-coding RNAs, overlapping transcripts and RNA elements that regulate transcription, such as riboswitches. Furthermore, processing and degradation of RNA has gained interest, because it has a significant impact on the composition of the transcriptome. In this review, we summarize recent findings and put them into a broader context with respect to the complexity of bacterial transcriptomes and its putative biological meanings.

Contents

1. Introduction 70
2. Transcriptional landscape 70
   2.1. Methods for transcriptome characterization 70
   2.2. Transcript types and their abundances 70
   2.3. The role of termination of transcription 70
3. Small RNAs 71
4. Sense and antisense overlapping transcripts 71
   4.1. Alternative and internal TSSs 71
   4.2. Antisense RNAs 72
      4.2.1. Excludons 72
5. *Cis*-regulatory elements 72
   5.1. Attenuators 72
   5.2. Riboswitches 72
   5.3. RNA thermometers 72
   5.4. Actuatons 72
6. Processing and degradation 73
   6.1. The manifold roles of RNAses 73
   6.2. sRNA-regulated mRNA decay 74
   6.3. Specific RNA processing and stabilization 74
7. Evolutionary conservation 75
8. Conclusion 76
   Outlook 76
   Acknowledgement 76
   References 76

* Corresponding author.
E-mail addresses: damir.stazic@biologie.uni-freiburg.de (D. Stazic), bjoern.voss@biologie.uni-freiburg.de (B. Voß).

http://dx.doi.org/10.1016/j.jbiotec.2015.09.041
0168-1656/

1. Introduction

The scientific community for a long time agreed on that the architecture of bacterial genomes is much simpler than that of eukaryotes. Concerning the structure and arrangement of genes, especially those coding for proteins, this is mainly explained by post-transcriptional processes. Due to splicing, eukaryotic genes are made up of introns and exons, where the latter become part of the mature mRNA. Alternative splicing and exon shuffling increase the variability of the genomic output. Importantly, also the non-protein-coding parts of the genome are often functional in the form of non-coding RNA. Major classes are microRNAs (Bartel, 2004), small interfering RNAs (Elbashir et al., 2001), Piwi interacting RNAs and long non-coding RNAs (review by Mercer and Mattick, 2013), some of which can be found even within introns of protein-coding genes. Besides this nesting of genes within other genes on the same strand, massive antisense transcription was reported (see review by Faghihi and Wahlestedt, 2009). Researchers readily recognized that these complex arrangements are a means to compensate for the surprisingly small number of genes in eukaryotic genomes, which is not sufficient to explain organism diversity. The current consensus is that RNA is at the heart of organism complexity (Licatalosi and Darnell, 2010). For bacteria, the picture is a bit different. Due to the lack of splicing and other post-transcriptional mechanisms, bacteria are still thought to have a much simpler way of encoding, transcribing and inheriting functions within their genomes. The availability of RNA-seq (Wang et al., 2009) and variants thereof, made in-depth studies of bacterial transcriptomes on nucleotide resolution possible, and shed light on an underestimated layer of activity, regulation and diversity (Georg et al., 2009; Sharma et al., 2010; Wurtzel et al., 2010, 2012; Mitschke et al., 2011; Guell et al., 2009; Georg and Hess, 2011).

In this review, we aim to comprehensively summarize the findings of the recent years, to put them into context to each other and to discuss the conclusions that can be drawn from these findings.

2. Transcriptional landscape

2.1. Methods for transcriptome characterization

The focus of this review is not on the advantages and disadvantages of different methods for studying bacterial transcriptomes, but nevertheless we want to provide a brief overview. The first large-scale studies on bacterial transcriptomes were based on tiling microarrays (Georg et al., 2009; Kumar et al., 2010; AbdelRahman et al., 2011; Gierga et al., 2012). Later, RNA-seq based studies became the method of choice (see review by Wang et al., 2009) and here especially those applying special treatments during library preparation. The most widely used approach among these is differential RNA-seq and its variants (see review in Sharma and Vogel, 2014). So far, more than 30 studies on more than 20 bacterial and archaeal species have been carried out and these numbers are steadily growing.

Central to the dRNA-seq method is the enrichment for primary transcripts, which is based on the tri-phosphate at their 5′-end. This makes them resistant to terminator exonuclease treatment or adaptor ligation, which is used to deplete or tag processing and degradation products that carry 5′-monophosphate or -hydroxyl groups. Libraries treated in such a way provide information on the native 5′-ends and thus on transcriptional start sites (TSSs). Commonly, a dRNA-seq study comprises also an untreated library, based on which the enrichment of reads at TSSs is inferred, and which is helpful for the identification of transcriptional units and operons (Sharma et al., 2010; Kopf et al., 2014; Pfeifer-Sancar et al., 2013).

2.2. Transcript types and their abundances

A major outcome of the aforementioned studies are genome-wide TSS maps. The identification of TSSs is often still performed manually, but algorithmic approaches exist (Jorjani and Zavolan, 2014; Amman et al., 2014) and will become more prominent in the future. Commonly, the next step is to classify the TSSs according to their location with respect to annotated genomic features. Here, often simple distance based classification schemes are used, e.g., a TSS within 200nt upstream of an annotated ORF is assigned to this ORF. This scheme is rather simplistic and can result in misclassifications if no further information, e.g. from the untreated library, is used. This can be overcome by using sophisticated algorithms for the inference of transcriptional units. Some of these work with standard RNA-seq data, e.g., Rockhopper (McClure et al., 2013) and Parseq (Mirauta et al., 2014), while others make use of the special information provided by dRNA-seq, e.g. RNAseg (Bischler et al., 2014).

When comparing results from different studies, it turns out that also the TSS classes themselves are not universal, and different names and classification schemes are used. For example, Sharma et al. (2010) term a TSS associated to an annotated gene pTSS or sTSS, which stands for primary and secondary TSS, respectively. TSSs associated to no known genomic feature are termed orphan (oTSSs), TSSs within genes internal (iTSSs) and in antisense orientation asTSSs. Mitschke et al. (2011) introduced a naming scheme that focusses on the resulting transcripts where the previously introduced pTSS and sTSS are both termed gTSS (TSS of a gene), oTSSs are called nTSSs, because they give rise to ncRNAs, and those giving rise to antisense RNAs are called aTSSs (similar to asTSSs before). The meaning of iTSS is retained. We adopt this notation which is also depicted in Fig. 1. Execpt for nTSSs the naming is not unambiguous, e.g. a transcript of a gene that starts internal to the previous gene would be a giTSS. This is fine for descriptive purposes but confusing for counting. A workaround is to apply a rule of preference for counting. Here, we use the rule gTSS >aTSS >iTSS. In Table 1 we provide a compilation of the numbers of transcript types that have been identified in selected dRNA-seq studies.

As a result of different library preparation protocols, differences in sequencing technology and depth, and different data analysis pipelines and TSS measures, the numbers in Table 1 cannot be compared directly. But, what these numbers show is that in most studied bacterial species large numbers of asRNAs and internal TSSs exist. Surprisingly, compared to previous studies the number of sRNAs (nTSSs) did not increase in the same order of magnitude as for asRNAs and iTSSs.

2.3. The role of termination of transcription

The transcript types introduced above put a focus on the initiation of transcription. Not less important is the termination of transcription, for which in general two mechanisms are known: Rho-dependent and Rho-independent termination. The process of termination in general adds another level of complexity to the transcriptome, because termination is not perfect and read-through is common. For adjacent genes this may result in fused read-through transcripts. This is also the case for Actuatons (see Section 5.4), with the additional peculiarity that the downstream gene lacks its own promoter. As a result, transcription of the downstream gene is controlled by the level of read-through and not directly by the activity of a promoter. Since termination is subject to regulation itself (Gusarov and Nudler, 2001) this makes complex expression patterns possible. Furthermore, many regulatory elements, such as Riboswitches and Attenuators, alter the transcriptional output