


[The Scientist](#) » [Magazine](#) » [Lab Tools](#)

Transcriptomics for the Animal Kingdom

Using RNA-seq to study nonmodel organisms

By Amy Maxmen | July 1, 2013

0 Comments Like 63 +1 2 [Link this](#) Stumble Tweet this



PROXY GENOME: The Glanville fritillary butterfly (*Melitaea cinxia*), which Christopher Wheat studies, is an important model species for studying population dynamics, specifically how variation in gene expression affects ecology.

© JIM MARDEN

efficiently analyze gene activity in organisms that don't have sequenced genomes, by building a transcriptome, which includes all of the RNA molecules produced by an organism—in all of its tissues and across its major life stages. Chopped into bits during RNA preparation, each sequenced fragment is like a piece of a puzzle, with the transcriptome serving as the completed image.

What makes transcriptomes so attractive to researchers studying nonmodel organisms is that they tend to be cheaper to build than genomes. What's more, when the work is done properly, the outcome can be more reliable than sequencing a genome, since some animals have gigantic genomes that are nearly impossible to assemble accurately.

In the past, researchers were limited to studying "about 20 out of 10 million species," says [Casey Dunn](#), an evolutionary biologist at Brown University in Providence, Rhode Island. "Now we can learn about the properties of all kinds of organisms in a way we could not do before." But, like any new and powerful technique, transcriptomics comes with a risk of misuse and misinterpretation. "It's a really exciting time to be a zoologist," Dunn says, "but it's still the Wild West." *The Scientist* talked with experts about how to build the most informative transcriptome, and asked for tips on how to make experiments more meaningful.

SETTING UP SHOP

Aim for fresh

When it comes to RNA experiments, fresh tissue is best. If the sample must be stored, snap freeze it in liquid nitrogen, or place it immediately in an Eppendorf tube filled with the RNA extraction buffer that comes in RNA purification kits from vendors such as [Invitrogen](#). Alternatively, some researchers store tissue in the stabilization reagent RNeasy (sold by a number of vendors including [Qiagen](#) and [Invitrogen](#)), and freeze it at -80 °C. But no kit can compensate for heavily damaged RNA, which degrades more easily than DNA.

"I've seen multiple studies where people say they sent tissue to collaborators at room temperature in the

If DNA is the architect's blueprint, then RNA is the contractor that determines what actually gets built in a cell. Gene-expression analyses, as measured by RNA levels, can reveal why two genetically similar species, such as chimpanzees and humans, differ so much. Within an individual, gene expression determines where a pinky forms, as opposed to a thumb. For years, biologists had been able to assess RNA levels more easily in model organisms whose genomes had been sequenced, and researchers who did not focus on model organisms were mainly left in the dark.

Rather than ask which genes within their species' genomes turn on at a particular time or in response to a particular environmental change, biologists studying nonmodel organisms were forced to fish for genes already identified in model animals. For example, a centipede researcher could only answer with certainty whether the many-legged creature activated the same genes as *D. melanogaster* as it grew its antennae.

With the advent of high-throughput, next-generation sequencing of RNA, called RNA-seq, this obstacle has largely been overcome. The technology allows zoologists to affordably and

Follow The Scientist



Subscribe!

Print or Digital

- iPad
- Kindle
- Tablet

Stay Connected with The Scientist

- The Scientist Magazine
- The Scientist Careers
- Neuroscience Research Techniques
- Genetic Research Techniques
- Cell Culture Techniques
- Microbiology and Immunology
- Cancer Research and Technology

Popular Posts

1. [Opinion: A Wolf in Sheep's Clothing](#)
2. [Genetic Test for Autism Refuted](#)
3. [Post-Publication Peer Review Mainstreamed](#)
4. [Is Cannabis Really That Bad?](#)
5. [It's in the Genes](#)

Current Issue



[View the October 2013 contents.](#)

mail, which means there's a good chance that their work is based on a cruddy sample," Dunn says. "If you are going to all of the trouble to make a transcriptome, it's worth it to start with good material."

Set controls early

Commercial kits for RNA purification provide step-by-step directions for how to convert an RNA sample into cDNA so that it can be sequenced. Adding a control sample at this first step is a must. Use this control throughout the experiment, so that after you sequence the samples you can check for contamination.

Test for quality

After RNA purification, [Christopher Wheat](#), an evolutionary biologist at Stockholm University in Sweden, checks the quality of each sample with a bioanalyzer (he uses the [2100 Bioanalyzer](#) from Agilent Technologies in Santa Clara, California). He looks for a normal distribution of fragment sizes over a range of 500 to 2,000 bases, which indicates that the RNA has not been degraded into low-quality, small fragments. "What you don't want to see is a peak around 100 base pairs, because that signals an accumulation of degraded RNA," he says.

PUTTING IT TOGETHER

Get more sequences

To generate a transcriptome, researchers assemble millions of RNA sequences into overlapping contiguous strands, called contigs, oriented in the correct direction. The first step is collecting the sequences of the extracted RNA. Two common next-generation sequencing platforms for RNA-seq are [Illumina](#) and [Roche 454](#). Many biologists currently prefer Illumina for transcriptome assembly because it yields more sequences per dollar, and like assembling a well-stocked library with a range of books in various editions, you want as many RNAs as possible in your transcriptome. However, many fragments will be from highly expressed, so-called housekeeping genes, which guide metabolism and other basic life-sustaining processes, so you need many reads in order to capture the sequences expressed at lower levels. "Half of the reads you'll get will hit 50 housekeeping genes," explains Wheat. Therefore, if you collect one million sequences, you're more likely to recover nonobvious genes among the other 500,000.

Decide when to stop

When [Cassandra Extavour](#), an evolutionary developmental geneticist at Harvard University, built a transcriptome for a crustacean commonly found near the sea, the amphipod *Parhyale hawaiiensis*, she accumulated as many sequences as she could. When is enough enough? Extavour answers this question by looking for a saturation point. If unique sequences continue to appear in abundance after she's analyzed a million reads, she will sequence another million. "Ideally you want to get all the RNA you can from every stage and every tissue," she says. "Remember, this is your proxy genome."

Size up the software

Most researchers agree that you shouldn't rely on genome-assembly software to build a transcriptome, but other than that rule, there is no single way to do it and the software changes frequently. Right now, many researchers use a free program called [Trinity](#), developed at the Broad Institute and the Hebrew University of Jerusalem. Trinity assembles transcriptomes using a three-step process that involves constructing full-length transcripts from fragments, clustering these transcripts into groups according to their likeness, and then teasing them apart into paralogous genes that are structurally similar but not the same. Trinity requires about 1 gig of RAM for every million reads that need to be assembled. Another popular (and also free) assembly software set is Velvet/Oases. [Velvet](#) is a genome-assembly program, but when coupled with [Oases](#), the program can handle RNA-seq data, filter out degraded fragments, and reconstruct transcripts that belong to genes.

Streamline the process

Processing your data before and after using programs like Trinity to assemble your transcriptome can improve your efficiency. For example, [Agalma](#) is a free pipeline program developed by Dunn's laboratory that gets your data ready for assembly and pushes it through the different analysis steps automatically. Before it assembles the transcriptome with Trinity, the program cleans up the data by discarding low-quality sequences, and also removes ribosomal RNA, which tends to be expressed in abundance. With the noise removed, Trinity moves faster and requires less of your computer's memory. Once the transcriptome is made and a researcher wants to construct an evolutionary tree with the new data, [Agalma](#) can do that too.

Ensure you've got quality sequences

After the transcriptome is created, Wheat suggests that researchers assess its quality with the online gene repository called the Basic Local Alignment Search Tool, or [BLAST](#). The length of the contig sequence you enter should be approximately equal to the corresponding gene sequence or contig from related species. Spot check with several other contigs.



NEFARIOUS CNIDARIAN: Known for its painful sting, the Portuguese Man of War (*Physalia physalis*) is the only siphonophore that lives at the surface of the ocean, using its large float as a sail.

COURTESY OF CASEY DUNN

Subscribe to RSS feed

-  [All](#)
-  [The Nutshell](#)
-  [News & Opinion](#)
-  [Careers](#)

GETTING THE MOST OUT OF YOUR TRANSCRIPTOME

Once the transcriptome is assembled, short RNA reads from experiments looking at anything from the RNA used as centipedes build their antennae to the genes found in jumping frogs can be mapped onto it. Here, a variety of open-access RNA-seq alignment software can help with the task. Popular choices include [Bowtie](#) and [Cufflinks](#)—both of which may be downloaded for free. Most RNA-seq alignment programs also estimate the relative abundance of transcripts, which can be useful for experiments that compare gene-expression levels.

Go off the grid

Many of the existing programs, such as Bowtie, were created for researchers who want to match the short RNA fragments gathered through RNA-seq data to a genome rather than a transcriptome, which can create problems.

Extavour says when she applied the program to her studies on the amphipod crustacean, it suggested that a third of her RNA-seq data had no matches in the amphipod transcriptome she had made. Instead of throwing her data out, she asked the bioinformatics technician in her laboratory to create a new alignment program.

With it, the team found a home for 90 percent of the RNA-seq reads. "Our own thing ended up working much better than using these programs off-the-shelf," Extavour says. (Her tech, Victor Zeng, has since begun his own start-up company that provides bioinformatics analysis to biologists working with large sequence data sets.)

Compare apples with apples

One dire mistake is to compare the expression level of genes between two species, or between two different genes. The problem with such an experiment is that some genes and some organisms are more amenable to sequencing than others, and what looks like a significant biological difference is really an artifact of the technique. "Maybe in a few years we will understand sequencing biases," says Dunn, "but right now we don't." Remember: same gene, same species.

Be realistic

Avoid comparisons between genes that are expressed at exceedingly low levels. Even if you manage to capture rare genes in your transcriptome, their scarcity will make quantitative assessments difficult.

Set your criteria first

When looking at the expression levels of a gene, Extavour advises researchers to use published research to decide how much of a difference is meaningful before the experiment begins, if possible. While most model organisms have benchmark standards after years of analysis, these benchmarks may not apply to your unstudied organism, so you must find your own. "Think about what you'll be satisfied with in advance," she says. "Is a 5-fold difference enough, and why?" For example, if previous investigations found that blocking the expression of a gene for skin spots in a tree frog does not change the spot pattern until only 20 percent of the expressed protein remains, a 2-fold difference in the expression of that gene will be too low to be informative, but a 10-fold difference should register.

Replicate, replicate, replicate

Don't just compare one sample to another one time; do it repeatedly. For example, in 2011, Dunn and his collaborators assessed gene-expression differences within different tissues in a species of siphonophore, an otherworldly marine animal related to jellyfish. Specifically, they sampled parts of the siphonophore devoted to feeding and parts that help it swim from three different siphonophores, in order to replicate the experiment three times ([PLOS ONE 6:e22953, 2011](#)). "Think of these experiments like polling in an election: just surveying a small subset of people may not reveal how the population at large feels, so you need to poll a number of subsets," Dunn explains.

Tags

[transcriptome](#), [techniques](#), [RNA-seq](#), [RNA sequencing](#) and [nonmodel organisms](#)

0 Comments    Like  63   +1  2  Link this  Stumble  Tweet this

Add a Comment



You

Sign In with your LabX Media Group Passport to leave a comment



FRAGILE CREATURES: One of the many siphonophores that Casey Dunn studies, *Marrus orthocanna* is a very fragile animal that can grow up to several meters in length.

COURTESY OF CASEY DUNN