

Innovative mark-recapture experiment shows patterns of selection on transcript abundance in the wild

Matthew Josephson¹ and James Bull¹

¹University of Calgary

April 6, 2021

Abstract

A fundamental aspect of evolutionary biology is natural selection on trait variation. Classically, selection has been estimated primarily on external morphological traits such as beak size and coloration, or on easily-assayable physiological traits such as heat-tolerance. As technologies and methods improved, evolutionary biologists began examining selection on molecular traits such as protein sequences and cellular processes. In a From the Cover manuscript in this issue of Molecular Ecology, Ahmad et al. (2021) continue this trend by estimating parasite driven selection on the molecular trait of transcript abundance in a wild population of brown trout (*Salmo trutta*) by uniquely combining a mark-recapture experimental design with non-invasive RNA sampling. Using transcript abundance to estimate selection allows for many different traits (each unique gene's transcript counts) to be tested in a single experiment, providing the opportunity to examine trends in selection. Ahmad et al.(2021) find directional selection strength on transcript counts is generally low and normally distributed. Surprisingly, transcripts under non-linear selection showed a disruptive selection bias contradicting previous comparative studies and theoretical work. This highlights the importance of within-generation selection studies, where mechanisms may differ from longer time frames. Their manuscript also highlights the benefits of an improved 3' RNA sequencing technique to measure gene expression.

Innovative mark-recapture experiment shows patterns of selection on transcript abundance in the wild

A fundamental aspect of evolutionary biology is natural selection on trait variation. Classically, selection has been estimated primarily on external morphological traits such as beak size and coloration, or on easily-assayable physiological traits such as heat-tolerance. As technologies and methods improved, evolutionary biologists began examining selection on molecular traits such as protein sequences and cellular processes. In a From the Cover manuscript in this issue of Molecular Ecology, Ahmad *et al.* (2021) continue this trend by estimating parasite driven selection on the molecular trait of transcript abundance in a wild population of brown trout (*Salmo trutta*) by uniquely combining a mark-recapture experimental design with non-invasive RNA sampling. Using transcript abundance to estimate selection allows for many different traits (each unique gene's transcript counts) to be tested in a single experiment, providing the opportunity to examine trends in selection. Ahmad *et al.*(2021) find directional selection strength on transcript counts is generally low and normally distributed. Surprisingly, transcripts under non-linear selection showed a disruptive selection bias contradicting previous comparative studies and theoretical work. This highlights the importance of within-generation selection studies, where mechanisms may differ from longer time frames. Their manuscript also highlights the benefits of an improved 3' RNA sequencing technique to measure gene expression.

For their study Ahmad *et al.* (2021), perform a mark-recapture study on a brown trout population that has a high prevalence of the parasite *Tetracapsuloides bryosalmonae* which causes the temperature-dependent

proliferative kidney disease (PKD) in salmonid fishes. PKD is an economically and ecologically important disease which causes significant mortality in salmonid species when water temperatures reach over 15°C (Hedrick, MacConnell, & de Kinkelin, 1993). The mark-recapture experiment was performed by capturing and clipping the pelvic fin of 278 wild young-of-the-year fish, releasing them back into their native stream, followed by an extensive effort to recapture nearly all (150) survivors one month later (Figure 1). By timing the initial sampling to shortly follow mass *T. bryosalmonae* spore release, the authors aimed to capture the transcriptomic response of individuals to infection, with mortality during the one month period between samplings thought to be primarily due to PKD infection.

Estimating selection on transcript abundance in a wild environment is a departure from previous studies on evolution of gene regulation which have used comparative approaches on divergent taxa. Highlighting the complementarity of these approaches, Ahmad *et al.* (2021), show an agreement with previous estimates of selection in finding that directional selection strength on transcript counts is quite weak and normally distributed (Groen *et al.*, 2020; Siepielski, Dibattista, & Carlson, 2009). However, previous comparative work and theory suggests selection on gene expression regulation is primarily through stabilizing selection, while Ahmad *et al.* (2021) show that in their experiment the most common and strongest form of non-linear selection on transcript abundance was disruptive selection (reviewed by Bedford & Hartl, 2009; Gilad *et al.*, 2006). The implications of this deviation from previous work will require further investigation to understand. One explanation is that the disruptive selection bias may be common for within generation selection in response to disease state, while long term adaptation to a specific environment may more frequently act through stabilizing selection on gene regulation. Specifically, Ahmad *et al.* (2021) propose optimal strategies for PKD tolerance in brown trout may be either large and energetically-expensive responses that lead to successful suppression of *T. bryosalmonae*, or the conservation of energy but allowing disease-associated damage, with intermediate responses incurring costs without substantial benefits.

An interesting approach used in the study was to sample fin tissue for RNA isolation; the small fin biopsy should not significantly affect fish survival. This differs from the common approach to focus on gene expression in tissues of interest (i.e. kidney/liver in immune studies) which typically requires lethal sampling. Measurement of gene expression from fin tissue may fail to capture interesting gene expression changes in immune organs in response to infection but allows a fundamentally new experimental design. Consistent with this, none of the gene expression networks that predicted survival were related to immune function. This should not be seen as a limitation, but rather a strength of their approach. Being able to non-lethally sample for RNA allows for selection on transcript abundance to be directly measured by survival and may shed light on new processes important to survival in the wild. To highlight this, a key gene expression network associated with parasite load and negatively correlated to survival identified from this work involved upregulation of many cell cycle genes, which the authors argue were associated with response to *T. bryosalmonae* infection and were not part of a generalized stress response, as stress is associated with downregulation of cell cycle processes (Burgess, Rasouli, & Rogers, 2014). While many RNA-seq studies have wide ranging implications their ecological importance is often unclear. For example a family of genes may be upregulated in response to some environmental change, but how this effects survival or reproduction is typically not measured. The experimental design exemplified by Ahmad *et al.* (2021) reveals new insight into PKD disease pathology and highlights how to perform an RNA-seq experiment in an ecologically relevant context.

Advances in sequencing technologies and techniques has led to a rise in gene expression studies in ecology, with the sequencing of mRNA (RNA-seq) becoming the most popular method for measuring gene expression. Traditional RNA-seq sequences entire mRNA transcripts, which gives splice variant information. However for most ecological applications, researchers are interested primarily in the amount of each gene being transcribed, and not concerned about splice variants. Obtaining full mRNA sequences thus represents wasted sequencing effort. A newly developed method termed TagSeq focuses only on the 3' end of transcripts, which in combination with a reference genome, is sufficiently informative to identify individual genes, thus increasing the number of individuals that can be reliably sequenced although at the cost of losing splice information (Meyer, Aglyamova, & Matz, 2011). Briefly, this method involves isolating total RNA, fragmentation, and cDNA synthesis with a poly-dT oligo to focus primarily on the 3' end of transcripts. This protocol has

been demonstrated to be more accurate and replicable than traditional NEBNext® RNA libraries (Lohman, Weber, & Bolnick, 2016). The reduced costs and increased sampling allowed by TagSeq represent a step forward for ecologists who are interested in gene expression quantification, and Ahmad *et al.* (2021) implement an optimized and commercially available version of TagSeq library preparation to measure transcript counts (Moll, Ante, Seitz, & Reda, 2014).

Overall, using improved RNA-seq methods to estimate within generation selection on transcript abundance in the wild demonstrates progress made in molecular ecology and the power of non-invasive sampling. The ability to cheaply and reliably measure transcript abundance in a large number of individuals should encourage more researchers to use transcript abundance as a measurable trait. Opening this avenue should lead to new insights on how natural selection pressure acts on gene expression regulation.

Ahmad, F., Debes, P. V., Nousiainen, I., Kahar, S., Pukk, L., Gross, R., ... Vasemägi, A. (2021). The strength and form of natural selection on transcript abundance in the wild. *Molecular Ecology*, 21:XXX-XXX doi: 10.1111/mec.15743

Bedford, T., & Hartl, D. L. (2009). Optimization of gene expression by natural selection. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4), 1133–1138. doi: 10.1073/pnas.0812009106

Burgess, A., Rasouli, M., & Rogers, S. (2014). Stressing mitosis to death. *Frontiers in Oncology*, 4 JUN(June), 1–7. doi: 10.3389/fonc.2014.00140

Gilad, Y., Oshlack, A., & Rifkin, S. A. (2006). Natural selection on gene expression. *Trends in Genetics*, 22(8), 456–461. doi: 10.1016/j.tig.2006.06.002

Groen, S. C., Čalić, I., Joly-Lopez, Z., Platts, A. E., Choi, J. Y., Natividad, M., ... Purugganan, M. D. (2020). The strength and pattern of natural selection on gene expression in rice. *Nature*, 578(7796), 572–576. doi: 10.1038/s41586-020-1997-2

Hedrick, R. P., MacConnell, E., & de Kinkelin, P. (1993). Proliferative kidney disease of salmonid fish. *Annual Review of Fish Diseases*, 3(C), 277–290. doi: 10.1016/0959-8030(93)90039-E

Lohman, B. K., Weber, J. N., & Bolnick, D. I. (2016). Evaluation of TagSeq, a reliable low-cost alternative for RNAseq. *Molecular Ecology Resources*, 16(6), 1315–1321. doi: 10.1111/1755-0998.12529

Meyer, E., Aglyamova, G. V., & Matz, M. V. (2011). Profiling gene expression responses of coral larvae (*Acropora millepora*) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. *Molecular Ecology*, 20(17), 3599–3616. doi: 10.1111/j.1365-294X.2011.05205.x

Moll, P., Ante, M., Seitz, A., & Reda, T. (2014). QuantSeq 3' mRNA sequencing for RNA quantification. *Nature Methods*, 11(12), i–iii. doi: 10.1038/nmeth.f.376

Siepielski, A. M., Dibattista, J. D., & Carlson, S. M. (2009). It's about time: The temporal dynamics of phenotypic selection in the wild. *Ecology Letters*, 12(11), 1261–1276. doi: 10.1111/j.1461-0248.2009.01381.x

Figure 1:

Selection on transcript abundance measured through a mark-recapture experiment. A) Non-lethal RNA sampling followed by 1 month of in-river selection. Recapture of all individuals required extensive effort but allowed for identification of all survivors and measurement of selection. B) DNA from post selection (September) used to Identify survivors. Selection differential estimated from transcript count in original (August) fin clip, and survivorship. Shown here is an example of a transcript under positive directional selection. This process was repeated by Ahmad et al., for all genes to analyze patterns of selection on transcript count in a wild environment.

