

SHORT COMMUNICATION

The activity level of a microbial community function can be predicted from its metatranscriptome

Damian E Helbling¹, Martin Ackermann^{1,2}, Kathrin Fenner^{1,2}, Hans-Peter E Kohler¹ and David R Johnson^{1,2}

¹Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland and

²Department of Environmental Sciences (D-UWIS), ETH Zürich, Zürich, Switzerland

The objective of this work was to improve our understanding of the quantitative predictive capabilities of metatranscriptomics. To meet this objective, we investigated whether we can predict the activity level of a specific biochemical function based on the abundance of the corresponding gene transcript within measured community metatranscriptomes. In addition, we investigated the lower limit of a microorganism's abundance that still allows detection of its transcripts within a metatranscriptome and prediction of the activity levels of the enzyme encoded by the transcript. To do this, we amended an undefined microbial community with varying fractions of an *Escherichia coli* strain that can catalyze a specific transformation reaction for the herbicide atrazine. We observed a linear and proportional relationship between the activity level of the transformation reaction and the abundance of its associated encoding transcript down to an *E. coli* cell density of 0.05% of the population.

The ISME Journal (2012) 6, 902–904; doi:10.1038/ISMEJ.2011.158; published online 17 November 2011

Subject Category: integrated genomics and post-genomics approaches in microbial ecology

Keywords: environmental genomics; metagenomics; xenobiotic degradation

Metagenomics approaches are promising techniques to characterize the genetic content of a variety of microbial communities (Venter *et al.*, 2004; DeLong *et al.*, 2006; Strous *et al.*, 2006; Rusch *et al.*, 2007; Turnbaugh *et al.*, 2009; Qin *et al.*, 2010). Metagenomic characterizations provide a basis for predicting the functions of a microbial community based on inferences from the types of genes or expressed gene content (messenger RNA) detected within the community. Whereas DNA-based metagenomics attempt to characterize the structure and metabolic potential of the whole community as represented by the community genome (Eisen, 2007), messenger RNA-based metatranscriptomics provide more precise information on community functions by identifying only expressed gene transcripts, and thereby more accurately reflect the pool of enzymes that are active within the community. Metatranscriptomics is a relatively new technical development, but has recently been applied to characterize functions within a variety of microbial communities (Leininger *et al.*, 2006; Frias-Lopez *et al.*, 2008; Poretsky *et al.*, 2009; Shi *et al.*, 2009; Ettwig *et al.*, 2010; Villa-Costa *et al.*, 2010).

The predictive capabilities of metatranscriptomics have thus far only been validated qualitatively, where connections have been made between the presence of specific transcripts and observed community functions (García Martín *et al.*, 2006; Ettwig *et al.*, 2010; McCarren *et al.*, 2010). Questions remain regarding whether metatranscriptomic datasets can also be used to make quantitative predictions on the activity levels of specific functions within microbial communities. The objective of this work was to improve our understanding of the quantitative predictive capabilities of metatranscriptomics by investigating: (i) whether differences in the abundances of specific transcripts within the measured metatranscriptomes of different communities reflect expected differences in the activity levels of their associated encoded enzymes; and (ii) how abundant a microorganism must be within a microbial community to detect its transcripts within a community metatranscriptome and predict the activity levels of its encoded enzymes. We specifically sought to test the case when proportionality is expected between the abundance of a specific transcript and the activity level of its associated encoded enzyme.

To meet this objective, we amended an undefined microbial community derived from a pilot-scale wastewater treatment plant with varying fractions of a recombinant strain of *E. coli* that constitutively expresses atrazine chlorohydrolase (*atzA*), an enzyme that transforms the herbicide atrazine (De Souza *et al.*,

Correspondence: DR Johnson, Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf 8600, Switzerland.
E-mail: david.johnson@eawag.ch

Received 7 July 2011; revised 22 September 2011; accepted 26 September 2011; published online 17 November 2011

1995). Atrazine chlorohydrolase catalyzes the hydrolysis of atrazine into 2-hydroxy atrazine, a function that was not detected in the community before *E. coli* addition. Six bioreactors were established in this manner resulting in relative *E. coli* cell densities that ranged between 0.05 and 100% of the total community cell density. Atrazine was added to each bioreactor at initial concentrations of $0.46 \mu\text{M}$ ($100 \mu\text{g l}^{-1}$). Samples were taken from each bioreactor over a period of 8 h and were used to measure the activity level of atrazine chlorohydrolase, quantified as the formation rate constant of 2-hydroxy atrazine; a single sample was taken at 4 h to quantify the abundance of the *atzA* transcript within the community by reverse transcription-quantitative PCR (RT-qPCR) analyses and metatranscriptome sequencing of extracted messenger RNA. Metatranscriptome sequencing was performed using the Illumina Genome Analyzer IIX (Illumina, San Diego, CA, USA) as described in the Supporting Information. RT-qPCR was selected as a complementary tool for this work because, in principle, all *atzA* transcripts can be amplified and counted during qPCR whereas only a small fraction of the overall sample pool is sequenced by the Illumina platform.

Figure 1 displays the results of the measured atrazine chlorohydrolase activities (quantified as the formation rate constant of 2-hydroxy atrazine) and the measured *atzA* transcript abundances (quantified from the metatranscriptome and by RT-qPCR analysis) for each of the six bioreactors. We observed a linear and proportional relationship between the number of *atzA* transcripts and the formation rate constant of 2-hydroxy atrazine (hr^{-1}) over four orders of magnitude (see the Supplementary Information for statistical analysis of linearity and proportionality). Uncertainty in formation rate constants (shown as vertical error bars in Figure 1) was

estimated from the 5th and 95th quantiles of parameter density distribution obtained from Markov Chain Monte Carlo sampling as described in the Supporting Information. Even with unquantified uncertainties in the sequence abundances and considering the actual uncertainties in the estimated formation rate constants, non-zero quantitative measurements were still possible for the *E. coli* function (2-hydroxy atrazine concentrations were above the limit of quantification) and *atzA* transcripts at the lowest *E. coli* abundance (0.05% of total community).

As we considered only one specific biological function as a proof of principle, we aimed to generalize our results by considering the catalytic efficiency of the function that was investigated. The catalytic efficiency of an enzyme for a given substrate is expressed with a specificity constant, for example, the ratio of the catalytic constant (k_{cat}) and the half-saturation (Michaelis) coefficient (K_M). It is plausible that the utility of metatranscriptomics as a predictive tool for community function should be enhanced as catalytic efficiency increases. Enzymes with higher catalytic efficiencies are more likely to have higher activities per unit transcript than enzymes with lower catalytic efficiencies. Consequently, for enzymes with high catalytic efficiencies, the detection of only a few transcripts makes it more likely that the associated enzyme function will also be measurable. The reported catalytic efficiency for the conversion of atrazine by atrazine chlorohydrolase is $7.4 \times 10^4 \text{M}^{-1} \text{s}^{-1}$ (De Souza *et al.*, 1996) and is among the lowest value of catalytic efficiency for other enzyme–substrate pairs in the literature (Snider *et al.*, 2004). Thus, our use of the atrazine chlorohydrolase encoded by the *atzA* transcripts was conservative for this proof of principle. As we could link *atzA* transcript levels with the activity levels of

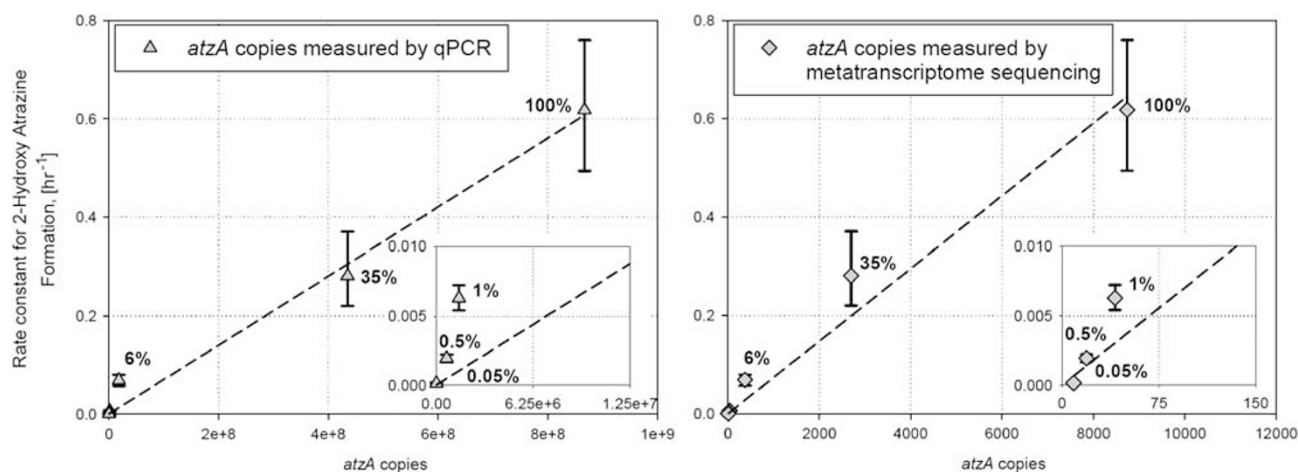


Figure 1 *atzA* gene transcripts detected by (▲) qPCR and (◆) metatranscriptome sequencing versus the rate of 2-hydroxy atrazine formation for each of the six *E. coli* titration levels (given as percentage of the total cell population). Error bars in the formation rate represent the 90% confidence interval from parameter estimation as described in the Supporting Information. The dashed lines are fit to the data based on model 1 linear regression by minimizing the sum of the squared residual errors and forcing a zero intercept. The coefficients of determination (r^2) for the model fits were 0.987 for the qPCR data and 0.970 for the metatranscriptome sequencing data. The inset plots show the data for the reactors with low *E. coli* cell densities plotted on the same axes (with a reduced scale) as the main plots.

atrazine chlorohydrolase (down to 9 *atzA* transcripts out of 7.3 million total sequence reads), we expect that it is possible to establish links between transcript levels and the activity levels of many other enzymes, even when only a few of these transcripts are detected. The effects of variability in post-transcriptional regulation were not directly investigated in this work, but these effects should not affect the relative abundance of active enzymes compared with transcript abundances. Thus, a proportional relationship between enzyme activity levels and transcript abundance should still be measurable.

In this work, we aimed to improve our understanding of the predictive capabilities of metatranscriptomics. We have shown that for the specific case of atrazine transformation catalyzed by atrazine chlorohydrolase, we could measure the expected proportional relationship between activity levels of atrazine transformation and the abundance of its associated encoding *atzA* transcript from community metatranscriptomes. Based on literature-reported values of catalytic efficiency, we further conclude that for many well-known enzyme-catalyzed reactions, if a transcript is detected at any level using currently available sequencing technologies then a quantitative prediction about the activity level of its encoded enzyme is plausible when the relationship between transcript abundance and activity is known. Moreover, our results demonstrate that metatranscriptomics approaches are not only useful for predicting the functions of abundant community members, but are also useful for predicting the functions of low-abundance community members that comprise as little as 0.05% of the total cell population. In summary, our results improve our understanding of the quantitative predictive power of metatranscriptomics and are thus relevant in a variety of sectors, including environment, biotechnology, agriculture, and human health and disease.

Acknowledgements

We thank Drs Joonhong Park, Helmut Burgmann, Jan Roelof van der Meer, Robert Nerenberg and Pilar Junier for critically reading the preliminary versions of this manuscript. We would like to additionally thank Drs Larry Wackett and Amit Aggerwal for providing the *E. coli* strain. This work was funded internally by Eawag Discretionary Funds for Research - Category: Seed Projects.

References

- De Souza ML, Sadowsky MJ, Wackett LP. (1996). Atrazine chlorohydrolase from *Pseudomonas* sp. strain ADP: gene sequence, enzyme purification, and protein characterization. *J Bacteriol* **178**: 4894–4900.
- De Souza ML, Wackett LP, Boundy-Mills KL, Mandelbaum RT, Sadowsky MJ. (1995). Cloning, characterization, and expression of a gene region from *Pseudomonas* sp.

- strain ADP involved in the dechlorination of atrazine. *Appl Environ Microbiol* **61**: 3373–3378.
- DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU *et al.* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.
- Eisen JA. (2007). Environmental shotgun sequencing: its potential and challenges for studying the hidden world of microbes. *PLoS Biol* **5**: 0384–0388.
- Ettwig KF, Butler MK, Le Paslier D, Pelletier E, Mangenot S, Kuypers MM *et al.* (2010). Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* **464**: 543–548.
- Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al.* (2008). Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* **105**: 3805–3810.
- García Martín H, Ivanova N, Kunin V, Warnecke F, Barry KW, McHardy AC *et al.* (2006). Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat Biotechnol* **24**: 1263–1269.
- Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW *et al.* (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**: 806–809.
- McCarren J, Becker JW, Repeta DJ, Shi Y, Young CR, Malmstrom RR *et al.* (2010). Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proc Natl Acad Sci USA* **107**: 16420–16427.
- Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP, Moran MA. (2009). Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environ Microbiol* **11**: 1358–1375.
- Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C *et al.* (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**: 59–65.
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yoeseff S *et al.* (2007). The Sorcerer II global ocean sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5**: 0398–0431.
- Shi Y, Tyson GW, DeLong EF. (2009). Metatranscriptomics reveals unique microbial small RNAs in the oceans water column. *Nature* **459**: 266–269.
- Snider MG, Temple BS, Wolfenden R. (2004). The path to the transition state in enzyme reactions: a survey of catalytic efficiencies. *J Phys Org Chem* **17**: 586–591.
- Strous M, Pelletier E, Mangenot S, Rattei T, Lehner A, Taylor MW *et al.* (2006). Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440**: 790–794.
- Turnbaugh PJ, Hamady M, Yatsuneko T, Cantarel BL, Duncan A, Ley RE *et al.* (2009). A core gut microbiome in obese and lean twins. *Nature* **457**: 480–484.
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso sea. *Science* **304**: 66–74.
- Villa-Costa M, Rinta-Kanto JM, Sun S, Sharma S, Poretsky R, Moran MA. (2010). Transcriptomic analysis of a marine bacterial community with dimethylsulfoniopropionate. *ISME J* **4**: 1410–1420.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)