

Q&A with Xin Zhou, author of our insect “squishome” paper.

Scott Edmunds 

Published April 2, 2013

Citation

Edmunds, S. (n.d.). In *GigaBlog*. GigaBlog. <https://doi.org/10.59350/7v5qw-ab764>

Keywords

Biology, Biodiversity, Genomics, GigaScience, Insect
Feature Image

Copyright

Copyright © Scott Edmunds 2013. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Following from our [previous blog posting](#), here we profile and interview Dr Xin Zhou, lead author of our recent "squishome" insect goo metabarcoding [paper](#). This NGS (next generation sequencing)-based work has already generated a lot of interest (see [this write-up](#) in Wired and this [blog posting](#) for nice examples), and here Dr Zhou gives more insight into the potential for the technique in studying biodiversity, as well as some of the quirky findings his team made validating the technique behind their laboratory in China.

Dr Xin Zhou is Director of the Environmental Genomics research group at [BGI](#), and Director of the Bio-resource Bank of the [China National GeneBank](#). A biodiversity expert and Entomologist by training, Dr Zhou carried out his postgraduate studies and postdoctoral training at Rutgers University and University of Guelph, managing barcoding projects for the [International Barcode of Life](#). From assembling and curating barcode reference libraries for a number of aquatic insect groups, his work has moved to the development of sequencing based analytical pipelines for bulk insect samples at Guelph, and since moving to China in October 2010, at BGI.

How is this method of PCR-free metabarcoding an improvement on previous techniques?

XZ: In PCR-based metabarcoding approaches, various primer sets are used to amplify target DNA fragments, which almost always introduce taxonomic biases, such that some organisms are easier to be detected while others are constantly missing or under-represented. This artificial bias poses a serious problem to all biodiversity studies where species composition is of a concern. Our work is the first of its kind that shows analyzing natural biodiversity samples doesn't have to rely on PCR therefore bypassing the primer issue. In addition, our work demonstrates that the PCR-free pipeline may potentially reveal species abundance from the mixed arthropod sample, providing yet another piece of crucial information to ecologists alike.

What does this NGS-based technology bring to it over just manually barcoding the collected samples?

XZ: Significantly reduced time and labor in sample processing as well as overall cost to analyze bulk samples.

What can you actually do with this the data, and what potential new applications does this technique enable? If you can say, what are you planning to do with this technique next?

XZ: This paper is a proof of concept demonstrating that natural bulk samples can be analyzed using NGS without having to relying on PCR amplification. This is the first step towards empirical applications of the new methodology in ecological and biodiversity related researches. We demonstrate that this new pipeline CAN work while a few technical issues can be improved for its wide implementation, such as in mitochondrial enrichment and tissue preservation. While trying to improve these technical details, we plan on increasing diversity scales of arthropod samples by analyzing those collected in tropical regions and arrays of insect samples collected from real-world ecological sampling designs.

Why did you collect the specimens you validated it with behind your lab at BGI [based in the large urban city of Shenzhen], and were you surprised by the number of specimens you detected?

XZ: This was an advantage working in a subtropical region where biological samples are relatively easy to obtain. Although the sampling was not comprehensive in terms of intensity of traps and number of species, we were surprised to see what we managed to collect in the middle of a community township. The 2 sampling sites were very close to each other yet there were merely ~10% of the total species being shared between them. Also, the fact that only very few of our barcoded specimens received a sequence match from the [Barcode of Life Data Systems](#), the world's largest barcode reference database, suggests that much of China's arthropod fauna still remains as a mystery, at least from a molecular aspect. On top of all that, we thought it would be an interesting idea to present BGI's headquarters in a scientific publication for the first time, with its GPS coordinates recorded in a meta-database.

Does this example say anything useful about the biodiversity in Shenzhen and the area around the BGI HQ?

XZ: as stated above, although the 2 arthropod bulk samples represent typical fauna of a secondary forestry ecosystem in Southern China, the overlap between samples was minimum and much of the community was poorly understood both morphologically and molecularly. We believe there is an urgent need to improve our knowledge on China's arthropod fauna. And we will start from where we live. We have a plan to barcode and metabarcode insects and plants of the Shenzhen municipal area.

In the paper it was interesting you found a novel COI [[cytochrome oxidase subunit I](#) – a common mitochondrial marker used in barcoding studies] from a [Lepidoptera](#) species not found in the reference library. Can you say a little more on this example, and is this a potentially new species?

XZ: Based on the quality of the nucleotide sequences and overall coverage of the novel barcode, we tend to believe that this is a real taxon that was somehow not detected in our morphological and barcode examinations. However, given the protocols used in this work, it is not possible to identify the exact source of this novel COI sequence. We listed a few potential possible reasons, including gut content, small residual tissues in the bulk sample, extracellular DNA etc. This novel sequence doesn't get a sequence match in any existing barcode databases. But this is not a big surprise as we know that Chinese insect species are not well-sequenced. The ultra-deep sequencing capacity of the NGS platforms opens up a new prospective where we are now capable of revealing diversity of the even-smaller-things-that-run-the-world via detecting their molecules. This would not have been possible if we had to rely on the visual cues of these organisms. In some sense, the contribution of NGS technology to biodiversity research is equivalent to what microscopes did to microbiology.

On that subject, what potential does this technique have to help discover new species, and how much can you actually tell about them using it?

XZ: NGS technology creates an alternative way to analyze biodiversity pattern and its temporal and spatial variations by detecting molecular or genomic heterogeneity ([MOTUs](#)) in bulk environmental samples. However, to make sense of these molecular operational units, one would have to compare this sequence information to well-curated sequence databases that are tied to conventional biological species concepts. A good example of these databases is the [Barcode of Life Data Systems](#), where millions of barcode sequences are linked to voucher specimens. My feeling is that the construction of sequence reference databases will remain critical in future molecular/genomic biodiversity research as it is a crucial step to provide linkages to the classic school of organismal science. However, NGS cataloging of world biodiversity can be performed in parallel. As long as meta-data are maintained for the bulk samples, biodiversity can be registered as MOTUs at first in a much accelerated fashion, and then be compared against existing reference databases available at the time. Known and (potentially) new species can be gradually revealed during this procedure. As biodiversity registration can be significantly accelerated using NGS, understanding biodiversity and especially interactions among species will be a long-term endeavor.

What are the implications for this technique in the growth of data taxa in the databases? As it is more high-throughput does it have the potential to massively increase the number of new entries?

XZ: This [PCR](#)-free approach can produce more accurate result in terms of species composition for bulk biological samples.

In terms of impact in increasing data entries in the databases, I believe this will be the future trend in biodiversity genomics. As the emergence of new technologies and rapid reduction in costs, the research community will be able to analyze much more biological samples in much shortened periods of time. The outcome will be an improved understanding of biodiversity changes based on consistent and standardized analysis procedures and intensified sampling (in terms of numbers of sampling sites across space and time and specimen numbers).

Is there anything else that you want to tell me about the technique?

XZ: the new PCR-free pipeline we created in this paper has further potentials in terms of construction reference genomes, such as mitochondrial and chloroplast genomes, in a much more economically efficient way. Based on our findings in the present work, much of the other mitochondrial genes of most of the insect species from the mixed sample can also be assembled with a decent [N50](#) value. For instance, the largest scaffold we managed to assemble from the insect soup was a moth representing almost the entire length of its mitochondrial genome. This means that with some tweak of the current pipeline, we would be able to sequence and assemble small genomes for many different species in one shot. Having a comprehensive reference library for mitochondrial genomes can solve many of the difficult questions faced in the classic barcoding community, such as primer designs for the standard barcode region for difficult groups, e.g., [Hymenoptera](#). Also this potential opens up the door to expanding classic barcoding methods from the current single-molecule approach to genomic screening.

Further reading:

1. Zhou X; et al., Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification *GigaScience* 2013 **2**:4 <http://dx.doi.org/10.1186/2047-217X-2-4>

The post [Q&A with Xin Zhou, author of our insect "squishome" paper](#). appeared first on GigaBlog.