Sequencing-Based Transcriptome-Wide Targeted Genotyping for Evolutionary and Ecological Studies

Evolutionary Bioinformatics
Volume 15: 1–3

DOI: 10.1177/1176934319836074

Xuan Zhu[1], Jing Wang[1,2], Jia Lv[1,2], Pingping Liu[1], Lingling Zhang[1,2], Wenqian Jiao[1], Cen Ma[1], Zhenmin Bao[1,3] and Shi Wang[1,2]

[1]MOE Key Laboratory of Marine Genetics and Breeding, College of Marine Life Sciences, Ocean University of China, Qingdao, China. [2]Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China. [3]Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China.

ABSTRACT: Transcriptome-wide targeted genotyping is highly attractive for evolutionary and ecological studies but, until recently, accomplishing this goal presented a major technical barrier for the study of non-model organisms. Our group has recently developed a high-throughput targeted genotyping approach (called HD-Marker) based on the high specificity and accuracy of oligo extension-ligation assays that facilitates the design of assays tailored to meet specific genotyping needs. HD-Marker allows for targeted genotyping of over 10 000 genes in a single tube, with strikingly high capture rate (98%-99%) and genotyping accuracy (97%-99%). With the remarkable advantages of cost-effectiveness and flexibility, we envision that HD-Marker has broad application potential in evolutionary and ecological studies.

KEYWORDS: HD-Marker, transcriptome-wide targeted genotyping, molecular markers, non-model organism

RECEIVED: February 1, 2019. **ACCEPTED:** February 13, 2019.

TYPE: Commentary

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by the National Key Research and Development Program of China (2018YFC0310802), National Natural Science Foundation of China (U1706203), Major Basic Research Projects of Shandong Natural Science Foundation (ZR2018ZA0748), and Blue Life Breakthrough Program of LMBB of Qingdao National Laboratory for Marine Science and Technology (MS2018NO01).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: The authors had filed a patent application based on the described technique.

CORRESPONDING AUTHOR: Shi Wang, MOE Key Laboratory of Marine Genetics and Breeding, College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China. Email: swang@ouc.edu.cn

COMMENT ON: Lv J, Jiao W, Guo H, et al. HD-Marker: a highly multiplexed and flexible approach for targeted genotyping of more than 10,000 genes in a single-tube assay. *Genome Res*. 2018;28:1919-1930. doi:10.1101/gr.235820.118. PubMed PMID: 30409770; PubMed Central PMCID: PMC6280760. https://www.ncbi.nlm.nih.gov/pubmed/30409770.

Genetic variation profiling is at the heart of modern biology and enables the study of important evolutionary and ecological questions by characterizing mutations, genetic drift, natural selection, and other processes of evolution.[1] The central goal of evolutionary genetics is to identify genetic changes and molecular mechanisms underlying phenotypic diversity and to understand the evolutionary pressures under which phenotypic diversity evolves.[2,3] The advent of next-generation sequencing (NGS) technologies has made it possible to cost-efficiently profile a large amount of genetic variations in less-studied non-model organisms and has boosted the rapid development of various high-throughput genotyping-by-sequencing (GBS) methods.[4,5] Unfortunately, most of these GBS methods use restriction enzymes for reduction of genomic complexity and therefore are mostly suited for random/novel marker discovery and genotyping, but not for targeting specific genomic regions/loci that are of particular interest to researchers.

The use of transcriptome-wide targeted genotyping is broadly desirable in evolutionary and ecological studies. Gene-related markers are derived from the transcribed genomic regions and such 'functional' markers are advantageous for pinning down causal genes underlying phenotypic/physiological traits or adaptive evolution.[6–8] Although transcriptome sequencing enables the efficient discovery of gene-related markers, cost-effective, large-scale targeted genotyping of these markers in non-model organisms is still a great challenge. For example, array-based genotyping is a popular and viable option for human and model organisms, but building custom arrays remains highly expensive to be applied to non-model organisms, and fixed arrays also lack the flexibility for rearrangement of targeted loci.[9] Sequence capture-based approaches are more suited for targeting genomic regions rather than specific loci of interest and therefore not practical for the genotyping of specific loci.[10]

To address this issue, our group has recently developed a sequencing-based GoldenGate approach (called HD-Marker), which allows high-throughput targeted genotyping of user-defined markers with high flexibility in choice of multiplex levels and marker types.[11] Different from the original GoldenGate technology that is built on the BeadArray platform with low-to-moderate marker multiplexity,[12] our HD-Marker approach is specifically designed for high-throughput sequencing platforms. For HD-Marker, genotyping a locus requires the design of two locus-specific probes (LSPs), each with a 3′ locus-specific region that targets the flanking sequence of the locus of interest and a 5′ portion with a universal polymerase chain reaction (PCR) primer-binding sequence (Figure 1). For a small probe panel, probes can be column-synthesized, whereas array-synthesized probes allow remarkably high multiplexity (over 12 000 loci) at a very low cost of ~US$0.001 per base. The locus under study (eg single-nucleotide polymorphisms [SNPs], microsatellites, or insertions/deletions [indels]) is placed in

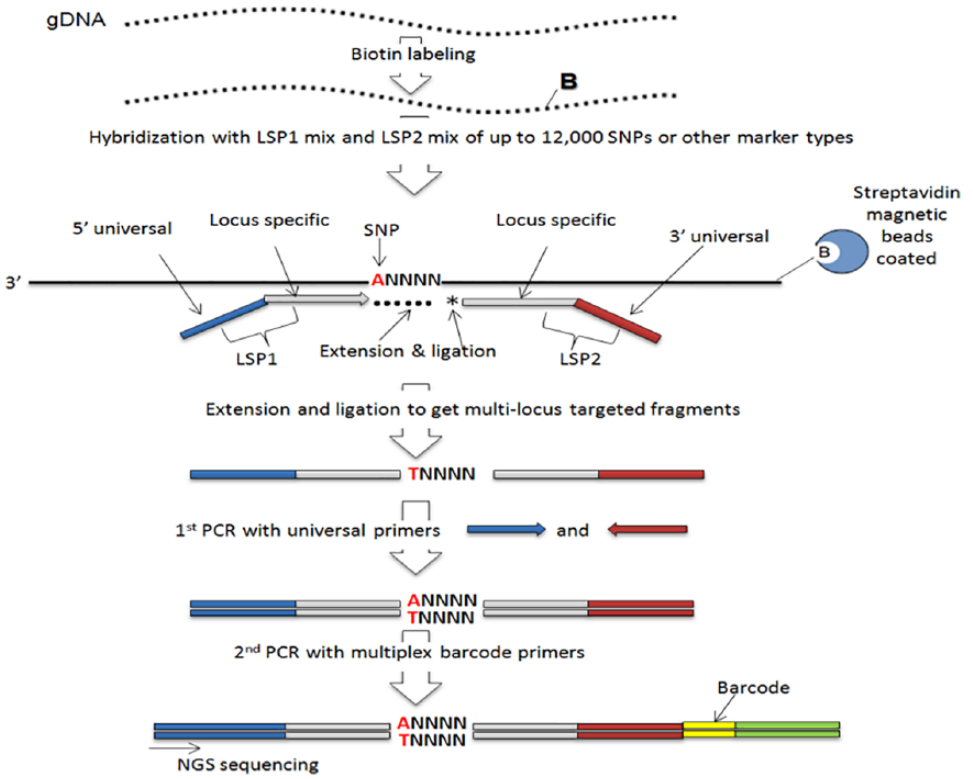

Figure 1. Schematic overview of the HD-Marker approach.

between the two LSPs, using DNA polymerase to fill in the gap and ligating the extended LSP to the downstream LSP, creating an amplifiable molecule. The preparation of an HD-Marker library starts with the attachment of genomic DNA to magnetic beads followed by the hybridization of two LSPs to the immobilized DNA via highly specific extension, ligation, and amplification. Libraries prepared from different samples can be pooled and sequenced on the NGS platform of choice. The whole procedure can be finished within 3 days, with all steps accomplished in a single tube.

We tested the applicability of HD-Marker using two non-model scallop species (*Chlamys farreri* and *Patinopecten yessoensis*) with abundant SNP and microsatellite marker resources. We evaluated four SNP multiplex levels (296, 795, 1293, and 12 472), with the first three (296, 795, and 1293) being column-synthesized and derived from the transcriptomic datasets of *C farreri*[13] and the last probe set (12 472) being array-synthesized and derived from the 11 771 genes of *P yessoensis*.[14] We showed high capture rate (98%-99%) and genotyping accuracy (97%-99%) on all multiplex levels, with remarkable evenness of allelic sampling. The feasibility of HD-Marker assay on targeting multi-gene families was further revealed, with high rates of loci detection (97.2%-100%), genotype calling (99.1%-100%), and genotyping accuracy (96.9%-100%). In addition, we demonstrated that HD-Marker can be used for targeted genotyping of non-SNP marker types (eg 50 microsatellites and 15 indels), with a genotyping accuracy of microsatellites (90%-100%) that is comparable to that of current methods (88%-98%).

Our HD-Marker approach has several major advantages, making it appealing for evolutionary and ecological studies. First, HD-Marker is built on the high-throughput NGS platform, which is much more cost-effective (as low as US$0.002 per genotype) than array-based platforms. Second, the use of NGS platform enhances the capacity and flexibility of HD-Marker, by removing the technical limitations of array formats. Third, a high-quality reference genome is not a prerequisite for applying HD-Marker and, for many non-model organisms, the widely accessible transcriptomic datasets and associated SNPs are already sufficient for the design of probes in HD-Marker. Fourth, HD-Marker allows the use of the same probe panel for cross-platform application. Fifth, library preparation of HD-Marker can be performed practically in any laboratory, as no costly instruments are required. Finally, the flexibility of HD-Marker allows the use of a variety of markers, such as SNPs, microsatellites, or indels, and of sequencing platforms, and is therefore easily adaptable to specific needs.

Transcriptome-wide targeted genotyping has long been appealing for evolutionary and ecological studies; however, for researchers studying non-model organisms, high cost and technical complexity make it unappealing. The HD-Marker approach retains the benefits of the GoldenGate assay but eliminates its array-based limitations by adopting the NGS

platform. HD-Marker enables researchers to create tailored assays to meet their specific genotyping needs, for example, focusing on targeted regions, candidate genes, or pathways. With the remarkable advantages of cost-effectiveness (as low as US$0.002 per genotype) and flexible choice of multiplex levels and marker types, we envision that HD-Marker would become a highly attractive tool with broad application potential in evolutionary and ecological studies.

Author Contributions

All authors contributed to the discussion and manuscript writing. XZ and JW contributed equally to this work.

REFERENCES

1. Rokas A, Abbot P. Harnessing genomics for evolutionary insights. *Trends Ecol Evol*. 2009;24:192–200. doi:10.1016/j.tree.2008.11.004.
2. Peichel CL, Marques DA. The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philos Trans R Soc Lond B Biol Sci*. 2017;372:20150486. doi:10.1098/rstb.2015.0486.
3. Romero IG, Ruvinsky I, Gilad Y. Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet*. 2012;13:505–516.
4. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet*. 2016;17:81–92.
5. Wang S, Lv J, Dou J, Lu Q_, Zhang L, Bao Z. Genotyping by sequencing and data analysis: RAD and 2b-RAD sequencing. In: Bioinformatics in Aquaculture (ed. Z Liu). Chichester, UK: John Wiley & Sons; 2017:338–355.
6. Andersen JR, Lübberstedt T. Functional markers in plants. *Trends Plant Sci*. 2003;8:554–560.
7. Jiao W, Fu X, Li J, et al. Large-scale development of gene-associated single-nucleotide polymorphism markers for molluscan population genomic, comparative genomic and genome-wide association studies. *DNA Res*. 2014;21:183–193. doi:10.1093/dnares/dst048.
8. De Wit P, Pespeni MH, Palumbi SR. SNP genotyping and population genomics from expressed sequences – current advances and future possibilities. *Mol Ecol*. 2015;24:2310–2323. doi:10.1111/mec.13165.
9. Thomson MJ. High-throughput SNP genotyping to accelerate crop improvement. *Plant Breed Biotech*. 2014;2:195–212.
10. Mertes F, Elsharawy A, Sauer S, et al. Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief Funct Genomics*. 2011;10:374–386. doi:10.1093/bfgp/elr033.
11. Lv J, Jiao W, Guo H, et al. HD-Marker: a highly multiplexed and flexible approach for targeted genotyping of more than 10,000 genes in a single-tube assay. *Genome Res*. 2018;28:1919–1930.
12. Fan JB, Oliphant A, Shen R, et al. Highly parallel SNP genotyping. *Cold Spring Harbor Symp Quant Biol*. 2003;68:69–78.
13. Hou R, Bao Z, Wang S, et al. Transcriptome sequencing and de novo analysis for *Yesso scallop (Patinopecten yessoensis)* using 454 GS FLX. *PLoS ONE*. 2011;6:e21560.
14. Wang S, Zhang J, Jiao W, et al. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol*. 2017;1:120. doi:10.1038/s41559-017-0120.