RESEARCH ARTICLE   **ARAŞTIRMA MAKALESİ**

A cost effective alternative method to ddRADseq library construction during size selection

ddRADseq kütüphanesi oluşturma işlemi fragman seçiminde uygun fiyatlı bir alternatif yöntem

Münevver Oral

Faculty of Fisheries and Aquatic Science, Recep Tayyip Erdoğan University, Rize, 53100, Türkiye   https://orcid.org/0000-0001-7318-6641

munevver.oral@erdogan.edu.tr

Abstract: Next generation sequencing (NGS) technologies constitute the most powerful scientific advance of 21st century with a promise of fast and cost effective data generation in biology. Yet, up to date NGS studies remain often limited to laboratories with established resources. In the present study, we employed construction of ddRADseq library by using routine lab consumables (agarose gel electrophoresis: AGE thereafter) compared to high-tech NGS consumables (paramagnetic beads) during size selection. The ddRADseq library was constructed for sequencing size selected based on universally used paramagnetic beads, while remaining aliquot was used as a template to assess the feasibility of ddRADseq library construction using AGE for labs with limited resources. Both libraries were optimised for 15 PCR cycles indicating similarity in template intensity. Post-PCR quantification of the libraries was comparable (~10 ng.µL^{-1}). Size distribution assessment revealed a cleaner pick at the ddRADseq library size selected manually based on AGE. Similarly, intercalating agent of Qubit confirmed the quantity of libraries was similar (>3 ng.µL^{-1}). Although being more time consuming due to pre-electrophoresis preparations, serial wash and staining steps, ddRADseq library construction is achievable using routine lab consumables provided to supply the adaptors and PCR primers for the initial wet-lab work. These results manifest the feasibility of ddRADseq library generation for labs with limited resources.

Keywords: Library preparation, ddRADseq lab workflow, next generation sequencing

Öz: Yeni nesil dizileme (YND) teknolojileri, biyolojide hızlı ve uygun maliyetli veri üretimi vaadi ile 21. yüzyılın en güçlü bilimsel ilerlemesini oluşturmaktadır. Yine de, güncel YND çalışmaları genellikle yerleşik kaynaklara sahip laboratuvarlarla sınırlı kalmaktadır. Bu çalışmada, kütüphane fragman seçimi sırasında yüksek teknoloji ürünü YND sarf malzemelerine (paramanyetik boncuklar) kıyasla rutin laboratuvar sarf malzemelerinden (agaroz jel elektroforezi: buradan itibaren AGE) kullanarak ddRADseq kütüphaneleri oluşturuldu. Standart ddRADseq kütüphanesi, evrensel olarak kullanılan paramanyetik boncuklara dayalı olarak seçilen fragmanlarla oluşturulurken, kalan kısım, sınırlı kaynaklara sahip laboratuvarlar için AGE kullanılarak aynı fragman büyüklüğünde ddRADseq kütüphanesi yapılabilirliğini değerlendirmek için bir şablon olarak kullanıldı. Her iki kütüphane de kalıp DNA yoğunluğunda benzerlik gösteren 15 PCR döngüsü için optimize edilmiştir. Kütüphanelerin PCR sonrası yoğunlukları benzerlik gösterdi (~10 ng.µL^{-1}). Boyut dağılımı değerlendirmesi, AGE ile manuel olarak seçilen ddRADseq kütüphane boyutunda daha temiz bir seçim olduğunu ortaya çıkardı. Benzer şekilde, Qubit ölçümleri de kütüphane DNA miktarının yakın olduğunu ortaya koydu (>3 ng.µL^{-1}). Elektroforez öncesi hazırlıklar, seri yıkama ve boyama adımları nedeniyle daha fazla zaman almasına rağmen, ddRADseq kütüphane kurulum işlemi başlangıç için gerekli adaptör ve PCR primerlerinin sağlanması kaydıyla rutin laboratuvar sarf malzemeleri kullanılarak gerçekleştirilebilir. Bu sonuçlar, sınırlı kaynaklara sahip laboratuvarlar için ddRADseq kütüphanesi oluşturmanın uygulanabilirliğini ortaya koymaktadır.

Anahtar kelimeler: Kütüphane hazırlama, ddRADseq laboratuvar iş akışı, yeni nesil dizileme

INTRODUCTION

The most recent breakthrough achieved in biological science is the development of next generation sequencing (NGS) technologies (Koboldt et al., 2013; McCombie et al., 2019; Hu et al., 2021). The ability of generating large number of genetic markers in relatively short period of time makes these technologies as a state of art methodology for genomic studies. NGS enables more individuals to be analysed at the same time by utilising high throughput sequencing, while massively parallel sequencing capacity multiplies the altitude of data generated with increasing accuracies as the depth of coverage rise. The validity of these technologies has widely been reviewed (MacLean et al., 2009; McCormack et al., 2013; Davey et al., 2011; Andrews et al., 2016; Fonseca et al., 2016;

Tan et al., 2019). However, regardless of their potential, NGS technologies are still limited to established laboratories in developed, high-income countries with large research budgets, thus also having related resources in terms of infrastructure and human power. This is essential as all stages of NGS including pre-library trials, library construction, sequencing and bioinformatics analysis require a demanding workload that necessitates cross-disciplinary collaboration (Knapp et al., 2015). Some pre-library steps (e.g. *in silico* analysis, adapter design) do not require extensive resources while sequencing is often shipped to public or private providers for the service. The bioinformatics analysis (e.g. quality check and filtering of data generated, variant calling, assembly, and downstream

genomic analysis) have often been put forward as the most critical part of the NGS workload (e.g. Guo et al., 2014; Shafer et al., 2016; Paris et al., 2017), however, few attention has been paid to wet-lab procedures in order to reduce the production cost and then to produce reliable and comparable libraries. Library preparation involves a series of molecular techniques which can be summarized in four main steps: (I) fragmentation of the genome of interest using restriction enzyme(s), (II) ligation of adaptors carrying sequencing primers for bridge amplification during sequencing by synthesis, (III) size selection ensuring desired fragments are captured, and (IV) enrichment of the library through PCR.

In the present study, we carried out a wet-lab work by using routine molecular genetic lab consumables (AGE) *versus* expensive NGS lab consumables (paramagnetic beads) during the size selection step of ddRADseq (double digest restriction-site associated DNA sequencing) library construction. We assessed two metrics as proxies while defining successful ddRADseq library construction procedures: (I) the number of PCR cycles required for enrichment of the library and (II) the quantification of the libraries as well as the distribution of the fragment range.

MATERIAL AND METHODS

Sampling

All individuals (*Salmo* spp.) considered in this study were sampled in the wild using pulsed DC electroshocking equipment under the supervision of Prof. Dr. Davut Turan following the Local Ethics Committee of RTE University for the use of animals in scientific experiments with a permit reference number of 2019/13. Specimens were released to nature once fin clips were taken. Trout species included *Salmo chilo*, *S. labecula*, *S. opimus*, *S. kottelati*, *S. platycephalus* and *S. tigridis* (Turan et al., 2011, 2012). Tissue samples of 55 individuals were used as a template to construct ddRADseq libraries.

Genomic DNA Extraction, Quantification and Quality Control (QC)

Genomic DNA was freshly extracted from individual fin clips using DNeasy® Blood & Tissue DNA extraction commercial kit (Qiagen, Valencia, CA), and performed on Qiacube DNA extraction robot (Qiagen, Valencia, CA). Following the manufacturer's guidelines, a RNase inhibition step was performed. These solutions constituted the stock DNA. The purity and the concentration of the extracted genomic DNA were initially assessed using NanoDrop spectrophotometry (2000C, Thermo Fisher Scientific, USA), while the integrity of the high molecular weight DNA was visualized on 0.8% agarose gel. Dilutions were made from stock DNA solutions down to 100 ng.µL^{-1} using double distilled water (ddH$_2$O). These served as working solutions throughout entire lab protocols. A final and more precise assessment of double stranded (ds)DNA concentration was carried out using Qubit fluorometer (Invitrogen, France) BR assay. Thus, based

on Qubit concentrations, samples were diluted down to 50 ng.µL^{-1} ready to use in ddRADseq library construction.

ddRADseq Library Construction

The ddRADseq library was generated originally by following Peterson et al. (2012) protocol with minor modifications detailed elsewhere (Palaiokostas et al., 2015; Leitwein et al., 2016; Oral et al., 2017). Each sample (200 ng DNA) was digested with two restriction enzymes, *Eco*RI-HF (G^AATTC) and *Msp*I (C^CGG) for 120 minutes at 37 °C in 25 µL reaction volume. The reactions were then treated with heat inactivation at 80 °C for 20 minutes. Some individuals were duplicated randomly in the library to reach 96 samples format so as to ensure higher coverage. P1 adaptor compatible with *Eco*RI and universally forked P2 adaptor compatible with *Msp*I were ligated to the digested DNA at 23°C room temperature for 120 minutes by adding 2 µL of each P1 and P2 adaptors (40 mM), 5 µL of ligase buffer, 1 µL of T4 DNA ligase (2K Units/µL) reaction volumes were made up to 25 µL using nuclease free water per sample (total reaction volume of 50 µL per sample). The ligation reaction was carried out in a thermal cycler, kept away from disruptions. Following heat inactivation at 65°C for 10 minutes, the plate carrying 96 samples was slowly cooled down to room temperature and pooled into a single ddRADseq library, labelled as *Pool*. The first purification was performed on ddRADseq library using 1X paramagnetic beads (Agencourt® AMPure® XP, Beckman Coulter, USA). Once purified, half of the reaction mix (25 µL) called DigLig (thereafter digested and ligated genomic DNA samples), was stored in -20 °C for *Pool* as a backup. Size selection (200-700 bp; see below for details) was carried out using paramagnetic beads (0.5-0.65X) by following manufacturer's guideline (the SPRI select User Guide, Beckman Coulter). The library was returned into individually labelled tube in 20 µL volume. Size selected library was then enriched by 15 cycles of PCR using 2x Phusion PCR master mix, 5 µL of library template, 2 µM of each PCR primer in 25 µL volume, reactions following cycling conditions of 98/65/72 °C for 10/30/45 seconds with a final extension at 72 °C for 5 minutes. PCR products were cleaned up using 0.75X paramagnetic beads to remove PCR primers. Purified and amplified PCR products were then quantified using both NanoDrop and Qubit. Library was assessed with Fragment Analyzer (Advanced Analytical Technologies, France) to determine DNA fragment size distribution. Based on this, one more round of 1X paramagnetic beads purification was carried out to remove primer dimers and small DNA fragments from the final library. After purification, the library was taken to the NGS sequence provider at minimum of 10 nM concentration in 20 µL volume. Library was sequenced (150 bp paired-end reads) on the Illumina HiSeq 2500 system.

Size Selection Paramagnetic Beads *versus* Agarose Gel

The final library (*Pool*) sent for sequencing was size selected based on paramagnetic beads. For that, an initial volume of 50 µL was used. First, larger fragments longer than

700 bp were removed from the library by using volume of low beads to ligation volume ratio (0.50X), then supernatant carrying smaller fragments was eliminated by using higher beads to sample volume (0.65X) (SPRIselect User Guide, 2012).

Alternatively, remaining DigLig was size selected using AGE on the same gel. For that, 1.1% agarose gel was poured using freshly made/diluted 0.5X TAE buffer with no ethidium bromide (EtBr). Once set, the gel was left on the fridge submerged in 0.5X TAE buffer for 2 hours. Several combs (7) were taped using autoclave tape to form a high volume comb that can accommodate 65 µL volume (55 µL Library and 10 µL 6X DNA loading dye ensuring 1X concentration on the gel) for the library (see Figure 1). Markers were loaded to both sides of the library so as to indicate marker cut regions. The electrophoresis tank was filled up with freshly made 0.5X TAE buffer that was used for the gel preparation previously and the tank was located on ice ensuring chilled run. The electrophoresis system was pre-run at a lower voltage (10V/cm) to ensure electronic contacts were sound. First, 10µL marker (customised) to both side of library was loaded. Then, 65 µL library DNA loading dye homogenous mix was loaded to gel slowly ensuring enough time interval for library to equilibrate in the well. Electrophoresis was initiated in low voltages (10 V/20 V/30 V) and gradually increased up to 90 V

to ensure no heating up in the tank buffer. The run was stopped when the loading dye migrated 3.5 cm from the origin (approx. 1 hr) (Figure 1A). The gel was placed on a glass and the library carrying fragment (ensuring loading dye band midway along the section) was safely separated from the gel using a sterile scalpel and stored in the fridge until required (Figure 1B-1C). Both sides of the gels were left behind (1-2 mm inside cut) so as to avoid the edge effect during electrophoresis. The remaining gel, carrying markers at both sides were stained in EtBr solution (2 µL of 5 mg/mL stock dissolved in 100 mL ddH$_2$O) for 5 minutes to visualise markers. Then, stained gel was washed with ddH$_2$O before come in contact with library carrying gel fragment (Figure 1C). Desired fragment size was identified under UV trans-illuminator located in a dark room and a small cut was marked just beside the markers that were previously left for avoiding electrophoresis edge effect (Figure 1D; orange box). Once both gels put together a horizontal cut was made and library carrying desired fragment size was taken from the agarose gel (Figure 1D; green box). Size selected library was first weighted and evenly split between three Eppendorf tubes to be purified using a column based gel extraction kit (Qiagen, France). The temperature of elution buffer was increased to 50 °C on heated block so as to increase the binding capacity. The remaining gel was visualised to obtain the restriction pattern of the library (Figure 1D).
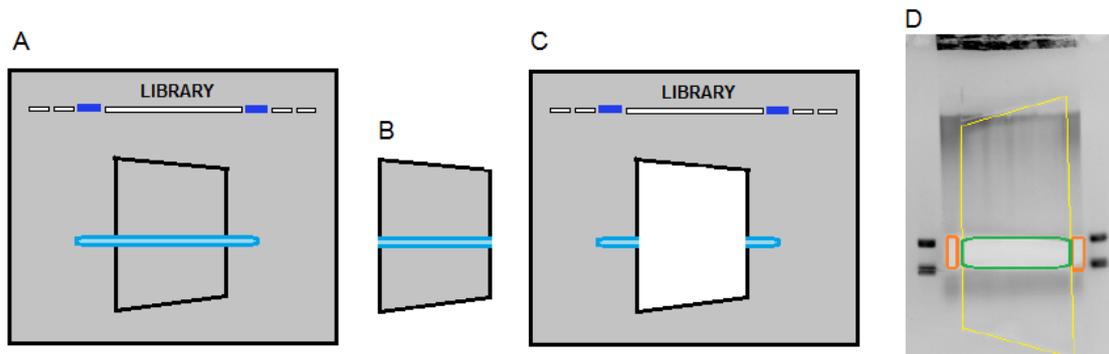

Figure 1. Schematic display of agarose gel based size selection of ddRADseq library. **A.** Electrophoresis was terminated once DNA loading dye migrated 3.5 cm from the origin; **B.** Library carrying gel was cut and stored at +4 °C; **C.** the remaining gel carrying markers was stained in EtBr solution, washed and a small cut was made under UV light indicating desired fragment size; **D.** ddRAD library gel processed throughout, yellow box: smear indicating restriction profile of the library as a positive control; orange box: indicative cut side of desired fragment size; green box: the fragment of interest ddRAD library

RESULTS

Quantification of ddRADseq libraries

One ddRADseq library, *Pool*, was constructed using the protocol size selected based on paramagnetic beads. The sequencing results and the downstream bioinformatics analysis of this library are discussed elsewhere (Oral et al., in preparation). Then, one more ddRADseq library was constructed, size selected based on agarose gel by using the remaining aliquot of DigLig genomic DNA of the same samples. This library was named as *AG_Pool* and was successfully extracted from the agarose gel. Size selection gel utilised was

thicker (>6 mm) than usual to ensure that the gel could compensate serial washing and staining during process. This is particularly the case for the size selection of multiple libraries. As a rule of thumb, a single gel in larger volume with gaps between the pools needs to be used so as to minimise any variation during electrophoresis. As temperature plays significant role during electrophoresis, the lower voltage was used so as to avoid warming up the buffer which leads to smiling effect on the gel as the heat increases. Therefore, the rationale behind the cold run was to minimise the diffusion of small fragments in the gel and obtain more precise sample fraction. The gel slice for *AG_Pool* weighted as 0,75 g and the

library was eluted in 50 µL volume (2x25 µL) using heated elution buffer.

Table 1 shows the quantification results of ddRADseq libraries generated. Size selected and PCR enriched ddRADseq libraries were initially quantified using NanoDrop. As the concentrations of the libraries were higher than 5 ng/µL (Table 1), 1:1 dilution was made using ddH$_2$O before checking the distribution of the DNA fragments with Fragment Analyzer. Therefore, Qubit accurately binds to dsDNA fragments resulted in lower concentrations for ddRADseq libraries produced following dilution. Similar library concentrations detected both in NanoDrop and Qubit were indicative of minimal variation achieved among the libraries generated regardless of the size selection method. Based on the intensity of libraries and concentrations, the number of PCR cycles could be decreased down to 12 or 13 cycles, which would also aid limiting the PCR duplicates.

Table 1. Quantification of the ddRADseq libraries

	PCR cycles	NanoDrop (ng/µL)	Qubit BR assay (ng/µL)	Average library size (bp)	Lib concentratio (nM)
Pool*	15	9.92	3.1	396	11.86
AG_Pool#	15	9.79	3.1	329	14.27

*: Paramagnetic beads based size selected and PCR enriched ddRADseq library
#: Agarose gel based size selected and PCR enriched ddRADseq library

Fragment distribution of ddRADseq libraries

In agreement with the quantifications of libraries by NanoDrop and Qubit, the Fragment Analyzer showed a similar distribution in the structure of the diagrams. The average sizes of the libraries were detected as 396 bp and 329 bp, respectively, in *Pool* and AG_*Pool* (Figure 2A and B), confirming that the desired fragment size of interest have been captured in both libraries. In this study, it should be further noticed that agarose gel-based size selection resulted in smaller size fragment distribution which was desired as dealing with larger fragments is more challenging in size selection with paramagnetic beads. A small pick observed consistently in *Pool* and AG_*Pool* is expected in species with duplicated genomes. These are likely results of repetitive regions represent redundant copies post-whole genome duplication of Salmonidae (e.g. Glasauer & Neuhauss, 2014). An additional round of 1X paramagnetic beads clean-up was carried out based on fragment analyser results so as to remove small fragments detected at 44 bp and 45 bp length in the diagrams, respectively (Figure 2A and B). Although detected in lower concentrations, if prominent, these can decrease the sequencing yield by competing to bind flow cell thus hamper proper cluster generation derived from adaptor ligated samples. The removal of this fragments following final 1X paramagnetic beads clean-up was confirmed by the results provided by the sequence provider prior sequencing.
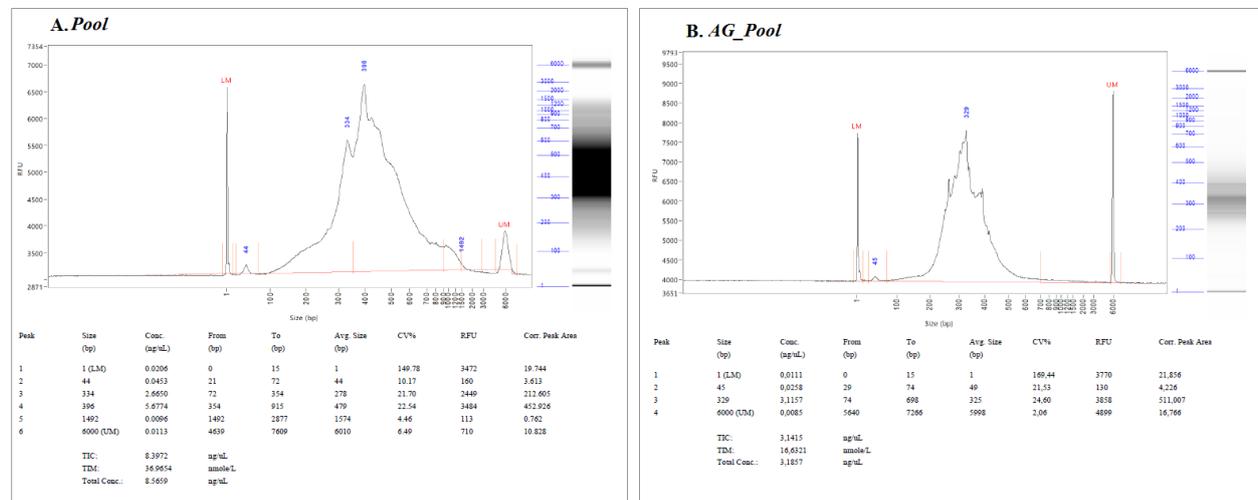

Figure 2. Distribution of the library fragments in ddRADseq libraries size selected based on (**A**) paramagnetic beads, *Pool* and (**B**) agarose gel electrophoresis AG_*Pool*

DISCUSSION

In the present study, we provided experimental data on the feasibility of ddRADseq library construction using cost effective routine molecular lab consumables. The rationale behind developing a cost-effective methodological improvement was motivated by the fact that feasibility of scaling up conventional molecular genetic laboratories to adapt working with cutting edge technologies such as ddRADseq.

PCR enrichment of the final library is essential step so as to ensure amplification of desired size range in the final library (Peterson et al., 2012). In standard ddRADseq libraries 12 to 18 PCR cycles are routinely used (Peterson et al., 2012; Capblancq et al., 2015; Palaiokostas et al. 2015; Yang et al., 2016; Burns et al., 2017; Oral et al., 2017; Cumer et al., 2021). The higher the PCR cycles increases the risk of PCR duplicates, while the lower PCR cycles camouflage the existing

diversity. Cumer et al. (2021) investigated the effect of some wet-lab procedures including DNA quantity and PCR cycles thus detected lower individual heterozygosity in 10 PCR cycles compared to optimal range of 15 PCR cycles both at the inter-species level for the animal model (in the butterfly species complex of *Coenonympha*) and at the intraspecific level for the plant model (in European/common beech, *Fagus sylvatica*) (Figure 2). Such debate indicates the significance of the trade-off exist between a satisfactory coverage and limitation of errors originates from the excess amount of PCR cycles (Hohenlohe et al., 2012). Alternatively, a series of test PCR can be set up for future ddRADseq library construction in half reaction volume (12.5 µL) by visualising the desired PCR cycle of the library on the agarose gel. These steps should be investigated in a case by case study.

Fragment Analyser results indicated similar size fragment ddRADseq libraries were successfully produced (Figure 2). While evaluated with another and less precise technology, a similar fragment size distribution (range: 300-350 bp) was also observed in ddRADseq libraries of Leitwein et al. (2016) on another trout species (*S. trutta*) (M. Leitwein & B. Guinand, *personal communication*).

Taken all together, agarose gel-based size selection provides a cost-effective alternative to expensive paramagnetic beads-based size selection. Agarose (0.70 € for 1 gr) is almost one quarter of the price compared to the paramagnetic beads (2.61 Euro for 100 µL) for the required amount. Additionally, given the availability of agarose as a routine molecular lab consumable and the experience that comes with it favours agarose as an economic alternative for any sized laboratory with limited research fund. Yang et al. (2016) suggested using conventional low melting agarose for size selection as opposed to the expensive automatized alternative of pippin prep in an experimental study to provide an alternative method for ddRADseq library construction for a wider community working on angiosperm plants. Similarly, final quality control of the ddRADseq library can efficiently be carried out on agarose gel to detect the average size of the library using an appropriate marker (e.g. 1kb GeneRuler). Based on the gel image, visualized on short (loading dye 1.5 cm away from the origin) and long run (loading dye 3.0 cm away from the origin) the minimum, maximum, mean and median size of the library fragment can be detected and this would be used for calculating the average size of the library. Final ddRADseq library concertation in nM can be then calculated using the following formula (provided from Illumina.com support web page) while gel image is sent to sequencing provider:

$$concentration\ in\ nM = \frac{\left(concentration\ in\ \frac{ng}{\mu l}\right)}{\left(660\frac{g}{mol} \times average\ library\ size\ in\ bp\right)} \times 10^6$$

The performance of our protocols was only evaluated based upon wet lab trials and experimental results. The biggest limitation of our study is to confirm the protocols by sequencing and provide the results from the data analysis (e.g. alignment rate to target reference genome, available SNP markers, basic statistics on sequence depth and coverage etc). However, the fact that all libraries provided sufficient requirements in terms of desired size range, concentration and available volume, we may anticipate the sequencing results would be of high quality and comparable between the two protocols (standard paramagnetic beads versus low-cost agarose gel electrophoresis).

In the present study, we provided experimental data on the feasibility of ddRADseq library construction using cost effective routine molecular lab consumables. As in all NGS experiments, the key to success is the availability of high molecular weight, intact genomic DNA and accurate quantification of dsDNA. Once this is achieved by using NanoDrop, Qubit and gel image, assuring a major band on the gel, sequencing produces sufficient amount of data as clearly demonstrated by Yang et al. (2016). In a recent study by Cumer et al. (2021) library construction pre-sequencing parameters including DNA quantity, number of PCR cycles during ddRADseq library preparation have shown to possess significant impact on the number of recovered reads and SNPs as well as on the number of unique alleles and individual heterozygosity. Additionally, same authors indicated the high reproducibility of the method provided to optimise the wet-lab procedures carefully. Furthermore, given the applicability of the protocol for any molecular laboratory, this study should motivate researchers in labs with limited resources to employ ddRADseq library construction provided to find partners that can supply adaptors and PCR primers for a start-up. Then, the cost only involves the investment of consumables for library preparation and sequencing. In the present study, expensive consumables were replaced by conventional alternatives where possible hence this protocol requires minimum technical investment for costly laboratory equipment and infrastructure. Therefore, taken all together, we anticipate our approach is applicable for any molecular laboratory with limited access to research funding thus the related human power.

CONCLUSION

In an effort to optimise the ddRADseq library construction for wider community, here we assessed the feasibility of size selection from agarose gel as opposed to paramagnetic beads. PCR conditions during enrichment of the libraries were identical for both groups, amplifying desired fragment size intensity. Besides, quantifications of the libraries based on NanoDrop and Qubit showed similar concentrations between two groups. Although agarose gel size selection was more laborious, this method produced a better fragment size distribution. Thus, given the availability of agarose and the experiences in electrophoresis this can be of a low-cost alternative to the high-tech paramagnetic beads during size selection of ddRADseq library construction for molecular laboratories with limited resources.

CONFLICTS OF INTEREST

The author declares no conflicts of interest.

DATA AVAILABILITY

Data generated within the course of the present study will be provided upon request to corresponding author.

ETHICS APPROVAL

This study was carried out with the approval of Local Ethics Committee of Recep Tayyip Erdoğan University for the use of animal in scientific experiments (Permit Number: 2019-13 and Date: 18.06.2019).

REFERENCES

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P.A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81–92. https://doi.org/10.1038/nrg.2015.28

Burns, M., Starrett, J., Derkarabetian, S., Richart, C. H., Cabrero, A., & Hedin, M. (2017). Comparative performance of double-digest RAD sequencing across divergent arachnid lineages. *Molecular Ecology Resources*, *17*(3), 418–430. https://doi.org/10.1111/1755-0998.12575

Capblancq, T., Després, L., Rioux, D., & Mavárez, J. (2015). Hybridization promotes speciation in *Coenonympha* butterflies. *Molecular Ecology*, *24*(24). https://doi.org/10.1111/mec.13479

Cumer, T., Pouchon, C., Boyer, F., Yannic, G., Rioux, D., and Bonin, A., & Capblancq, T. (2021). Double-digest RAD-sequencing: do pre- and post-sequencing protocol parameters impact biological results? *Molecular Genetics and Genomics*, *296*, 457–471. https://doi.org/10.1007/s00438-020-01756-9

Davey, J. W., Hohenlohe, P. A, Etter, P. D., Boones, J.Q., Catchen, J.M., & Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, *12*(7), 499–510. https://doi.org/10.1038/nrg3012

Fonseca, R.R., Albrechtsen, A., Themudo, G.E., Madriagal, J.R., Sibbesen, J.A., Maretty, L., Mendoza, M.L., Campos, P.F., Heller, R, & Pereira, R.J. (2016). Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, *30*, 3-13. https://doi.org/10.1016/j.margen.2016.04.012

Glasauer, S. M. K. & Neuhauss, S. C. F. (2014). Whole-genome duplication in teleost fishes and its evolutionary consequences. *Molecular Genetics and Genomics*, *289*(6), 1045–60. https://doi.org/10.1007/s00438-014-0889-2

Guo, Y., Ye, F., Sheng, Y., Sheng, Q., Clark, T., & Samuels, D.C. (2014). Three-stage quality control strategies for DNA re-sequencing data. *Briefings in Bioinformatics*, *15*(6), 879–889. https://doi.org/10.1093/bib/bbt069

Hohenlohe, P. A., Catchen, J., & Cresko, W. A. (2012). Population Genomic Analysis of Model and Nonmodel Organisms Using Sequenced RAD Tags. In *Data Production and Analysis in Population Genomics* (pp. 235–260). Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-61779-870-2_14

Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology 82*; 801–811.

Illumina.com. (Accession date: 20.05.2022; 9:00). Converting ng/µl to nM when calculating dsDNA library concentration. (https://emea.support.illumina.com/bulletins/2016/11/converting-ngl-to-nm-when-calculating-dsdna-library-concentration-.html?langsel)

Knapp, B., Bardenet, R., Bernabeu, M.O., & Deane, C.M. (2015). Ten simple rules for a successful cross-disciplinary collaboration. *Plos Computational Biology*, *11*(4), e1004214. https://doi.org/10.1371/journal.pcbi.1004214

Koboldt, D.C., Steinberg, K.M., Larson, D.E., Wilson, R.K., & Mardis, E.R. (2013). The next-generation sequencing revolution and its impact on genomics. *Cell*, *155*:27–38. https://doi.org/10.1016/j.cell.2013.09.006

Leitwein, M., Gagnaire, P. A., Desmarais, E., Guendouz, S., Rohmer, M., Berrebi, P., & Guinand, B. (2016). Genome-wide nucleotide diversity of hatchery-reared Atlantic and Mediterranean strains of brown trout *Salmo trutta* compared to wild Mediterranean populations. *Journal of Fish Biology, 89*, 2717–2734. https://doi.org/10.1111/jfb.13131

MacLean, D., Jonathan D.G.J., & Studholme, D. J. (2009). Application of next-generation sequencing technologies to microbial genetics. *Nature Reviews Microbiology*, 7, 287-296. https://doi.org/10.1038/nrmicro2088

McCombie, W.R., McPherson, J.D., & Mardis, E.R. (2019). Next Generation Sequencing Technologies, *Cold Spring Harb Perspect Med* 2019;9:a036798. https://doi.org/10.1101/cshperspect.a036798

McCormack, J.E., Hird, S.M., Zellmer, A. J., Carstens, B.C., & Brumfield, R.T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, *66*(2), 526-538. https://doi.org/10.1016/j.ympev.2011.12.007

Oral, M., Colléter, J., Bekaert, M., Taggart, J.B., Palaiokostas, C., McAndrews, B.J., Vandeputte, M., Chatain, B., Kuhl, H., Reinhard, R., Peruzzi, S., & Penman, D.J. (2017). Gene-centromere mapping in meiotic gynogenetic European seabass. *BMC Genomics, 18*, 449. https://doi.org/10.1186/s12864-017-3826-z

Palaiokostas, C., Bekaert, M., Khan, M. G. Q., Taggart, J.B., Gharbi, K., McAndrew, B.J., & Penman, D.J. (2015). A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). *BMC Genomics*, *16*(1), 171. https://doi.org/10.1186/s12864-015-1383-x

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: A road map for stacks. *Methods in Ecology and Evolution*, *8*(10), 1360-1373. https://doi.org/10.1111/2041-210X.12775

Peterson, B.K., Weber, J., Kay, E.H., Fisher, H.S., & Hoekstra, H.E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *Plos One*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C.W., & Wolf, J.B.W. (2016). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, *8*(8), 907–917. https://doi.org/10.1111/2041-210X.12700

SPRIselect User Guide. (Accession date: 17.06.2022; 14:20). *Size selection based of paramagnetic beads*. p. 1-8. (https://research.fhcrc.org/content/dam/stripe/hahn/methods/mol_biol/SPRIselect%20User%20Guide.pdf)

Tan, M.P., Wong, L.L., Razali, S.A. Aleng, N.A., Nor, S.A.M., Sung, Y.Y., Peer, Y.V., Sorgeloos, P., & Daniel, M.D. (2019). Applications of Next-Generation Sequencing Technologies and Computational Tools in Molecular Evolution and Aquatic Animals Conservation Studies: A Short

Review. *Evolutionary Bioinformatics*, *15*, 1-5. https://doi.org/10.1177/1176934319892284

Turan, D., Kottelat, M., & Bektaş, Y. (2011). *Salmo tigridis*, a new species of trout from Tigris River, Turkey (Teleostei: Salmonidae). *Zootaxa 2993*, 23–33. https://doi.org/10.11646/zootaxa.2993.1.2

Turan, D., Kottelat, M., & Engin, S. (2012). The trouts of the Mediterranean drainages of southern Anatolia, Turkey, with description of three new species (Teleostei: Salmonidae). *Ichthyological Exploration of Freshwaters*, *23*, 219–236.

Yang, G.Q., Chen, Y.M., Wang, J.P., Guo, C., Zhao, L., Wang, X.Y., Guo, Y., Li, L., Li, D.Z., & Guo, Z.H. (2016). Development of a universal and simplified ddRAD library preparation approach for SNP discovery and genotyping in angiosperm plants. *Plant Methods*, *12*, 39. https://doi.org/10.1186/s13007-016-0139-1