HHS Public Access

Author manuscript

J Microbiol Methods. Author manuscript; available in PMC 2023 December 26.

Published in final edited form as:

J Microbiol Methods. 2016 November ; 130: 23–26. doi:10.1016/j.mimet.2016.08.020.

Evaluation of library preparation methods for Illumina next generation sequencing of small amounts of DNA from foodborne parasites

Fernanda S. Nascimento[a], **Yuping Wei-Pridgeon**[a,b], **Michael J. Arrowood**[c], **Delynn Moss**[c], **Alexandre J. da Silva**[a,1], **Eldin Talundzic**[d,e], **Yvonne Qvarnstrom**[a,*]

[a] Parasitic Disease Branch, Division of Parasitic Diseases and Malaria, Center for Global Health, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA 30329, United States

[b] IHRC Inc, 2 Ravinia Drive, Suite 1750, Atlanta, GA 30346, United States

[c] Waterborne Disease Prevention Branch, Division of Foodborne, Waterborne, and Environmental Diseases, National Center for Enteric and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA, 30329, United States

[d] Malaria Branch, Division of Parasitic Diseases and Malaria, Center for Global Health, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA 30329, United States

[e] Atlanta Research and Education Foundation, 1670 Clairmont Road, Suite 5A104, Decatur, GA 30033, United States.

Abstract

Illumina library preparation methods for ultra-low input amounts were compared using genomic DNA from two foodborne parasites (*Angiostrongylus cantonensis* and *Cyclospora cayetanensis*) as examples. The Ovation Ultralow method resulted in libraries with the highest concentration and produced quality sequencing data, even when the input DNA was in the picogram range.

Keywords

Next generation sequencing; Illumina; MiSeq; Library preparation; *Angiostrongylus cantonensis* ; *Cyclospora cayetanensis*

Whole genome sequencing (WGS) is a promising technique to develop advanced diagnostic, molecular epidemiology, and source tracking tools for food-borne parasites of public health importance. A major bottleneck in genome sequencing is library construction, which refers to the preparation of the nucleic acid into a form that is compatible with the sequencing system to be used (Head et al. 2014). Most of the next generation sequencing platforms adhere to the same basic library production strategies, including DNA fragmentation, end

* Corresponding author. bvp2@cdc.gov (Y. Qvarnstrom).

[1]Present address: Office of Applied Research and Safety Assessment, Center for Food Safety and Nutrition, Food and Drug Administration, Laurel, MD 20708.

repairs and adapter ligation. Standard library preparation methods require large quantities of nucleic acids, making it challenging to apply to microorganisms that cannot be propagated in the laboratory (Bhattacharya et al. 2002, Chandra et al. 2014, Lay et al. 2010). However, recent advances make it possible to produce libraries with much lower amounts of input material (Parkinson et al. 2012).

Previous comparisons of commercial Illumina library preparation kits revealed that methods used influence sequencing results (Lan et al. 2015; Rhodes et al. 2014). To identify method(s) suitable for WGS of eukaryotic parasites, we compared four library preparation kits intended for low input DNA amounts: NEBNext Ultra DNA Library Prep kit (New England Biolabs Inc.), Ovation Ultralow Library System (Nugen Technologies Inc.), ThruPlex FD Prep kit (Rubicon Genomics Inc.) and Nextera XT DNA Library Kit (Illumina). The comparison was made with genomic DNA from *Angiostrongylus cantonensis*, a nematode associated with eosinophilic meningitis worldwide (Wang et al. 2012) but whose genome is not fully characterized (Yong et al. 2015; Morassutti et al. 2013). DNA was extracted from an *A. cantonensis* adult worm using DNeasy® Blood and Tissue Kit (QIAGEN) and quantified using Qubit dsDNA HS Assay (Invitrogen). One nanogram of DNA was used as starting material for each kit: intact genomic DNA for Nextera XT (since it employs enzymatic fragmentation); and mechanically-fragmented DNA, using conditions for 300 base pairs in an M220 Focused-Ultrasonicator™ (Covaris Inc.), for the other three methods. The quality and quantity of the libraries were assessed in a 2200 TapeStation (Agilent Technologies).

The Ovation, ThruPlex and NEBNext libraries had similar size distribution, but the Ovation library was considerably higher in concentration (Fig. 1a). The NEBNext library produced adapter dimers, as evidenced by the presence of a smaller peak beside the main library. The Nextera library could only be detected using a genomic DNA screen tape, revealing that most of the library consisted of very large fragments, indicating insufficient enzymatic fragmentation (Fig. 1b). Possible reasons for this are inadequate purity or composition of the parasite DNA. Inaccurate DNA quantification can reportedly lead to production of longer fragments due to an unfavorable ratio between the tagmentation enzyme and the number of DNA molecules accessible to the enzyme (Adey and Shendure 2012). Applying a size exclusion step can eliminate adapter dimers from the NEBNext library as well as the larger fragments from the Nextera library. However, size exclusion was not applied in order to maintain consistent standards for comparison of the different methods. Besides, size selection steps do not always remove long fragments in Nextera libraries (Kim et al. 2013; Lamble et al. 2013).

The libraries were sequenced using MiSeq Reagent Kit v2 (500 cycles) (Illumina). The quality of sequencing results depends both on the library quality (van Dijk et al. 2014) and the bioinformatics tools used to analyze the sequencing data, such as trimming and assembly algorithms (Ekblom and Wolf 2014). The quality of the sequence data obtained in this study was assessed using FastQC 0.11.4. The BBduk plugin in Geneious R9 (Biomatters Ltd.) was used for trimming. To ensure an unbiased comparison, the same number of trimmed reads was randomly extracted from each sequenced library and assembled using the Geneious *de novo* assembler and SPAdes assembler 3.5 (Bankevich et al. 2012). The

assemblies were compared using QUAST (Gurevich et al. 2013). Approximately the same fraction of reads from each library, *i.e.*, 78%, 79%, 83% and 73% from NEBNext, Nextera, Ovation and ThruPlex libraries, respectively, passed the quality filters. Overall, sequencing data from the Ovation library produced the best assemblies, with the highest N50 values and the longest contigs (Table 1). However, the Nextera library produced longer total contig length, indicating slightly better genome coverage. Thus, when compared to the other library preparation methods, Ovation yielded libraries with the highest concentration and resulted in good quality sequences that assembled into the fewest, longest *de novo* contigs.

The Ovation method was further evaluated using genomic DNA from *Cyclospora cayetanensis*, which causes gastroenteritis and is associated with large outbreaks (Abanyie et al. 2015). This parasite only infects humans and cannot be propagated in the laboratory. Usually less than 10^5 parasites are present in a typical diagnostic human fecal specimen: such specimens are expected to yield only picograms of parasite DNA. Oocysts were purified from human feces (Arrowood and Donaldson 1996), sorted and counted in a FACSAria III (BD Biosciences) flow cytometer. DNA was extracted from 100,000 and 10,000 oocysts following 15 freeze:thaw cycles (Lalonde et al. 2013) and subsequently purified with the DNAeasy kit (QIAGEN). The *Cyclospora* DNA concentration was below the detection limit of the Qubit HS Assay in both samples, so a real-time PCR was performed (Verweij et al. 2003) and the resulting Ct values were used to guide the number of amplification cycles required during library preparation. The libraries were sequenced and the resulting reads were quality trimmed as described above for the *Angiostrongylus* experiment. The *Cyclospora* libraries had a broad size distribution with a peak size of about 1700 bp, indicating inefficient DNA shearing (Fig. 2). They nevertheless produced acceptable sequencing results in the MiSeq: 77–75% of the raw reads passed the quality filter. Trimmed reads were mapped to the draft *C. cayetanensis* genome (Qvarnstrom et al. 2015) assembly using Bowtie2 (Langmead and Salzberg 2012) in Geneious® R9. About 97% of the estimated genome length (45 Mb) was recovered from each library, indicating no or very low bias during the library preparation.

In summary, all library methods evaluated produced libraries of sufficient concentration and resulted in similar quality of the raw reads. However, libraries produced using Ovation were superior in overall yield, even from undetectable DNA amounts, and produced acceptable sequencing assembly results. We therefore consider the Ovation method most suitable for WGS of ultralow amounts of parasite DNA.

This study was supported by the CDC's Advanced Molecular Detection and Response to Infectious Disease Outbreaks Initiative. Dr. Nascimento was supported by the Brazilian National Counsel of Technological and Scientific Development (CNPq) fellowship [236608/2013–4]. The Biotechnology Core Facility Branch in the CDC's National Center for Zoonotic and Emerging Infectious Diseases provided technical support. We thank Bernard Asuncion from the Vector Control Inspector, Hawaii Department of Health, Alessandra L. Morassutti from the Pontifícia Universidade Católica do Rio Grande do Sul, Brazil, and Lisa Rascoe, currently at the Department of Biomedical Sciences of Florida State University, for their technical assistance to obtain *A. cantonensis* from infected rats.

References

Abanyie F, Harvey RR, Harris J, Wiegand R, Gaul L, Desvignes-Kendrick M, Irvin K, Williams I, Hall R, Herwaldt B, Bosserman E, Qvarnstrom Y, Wise M, Cantu V, Cantey P, Bosch S, da Silva AJ, Hardin A, Bishop H, Wellman A, Beal J, Wilson N, Fiore AE, Tauxe R, Lance S, Slutsker L, Parise M, and the Multistate Cyclosporiasis Outbreak Investigation Team. 2013 multistate outbreaks of *Cyclospora cayetanensis* infections associated with fresh produce: focus on the Texas investigations. Epidemiol. Infect. 2015;143:3451–8 [PubMed: 25865140]

Adey A, Shendure J, 2012. Ultra-low-input, tagmentation-based whole genome bisulfite sequencing. Genome Res. 22, 1139–1143. [PubMed: 22466172]

Arrowood MJ, Donaldson K, 1996. Improved purification methods for calf-derived *Cryptosporidium parvum* oocysts using discontinuous sucrose and cesium chloride gradients. J. Eukaryot. Microbiol. 43, 89S. [PubMed: 8822880]

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA, 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. [PubMed: 22506599]

Bhattacharya S, Vijayalakshmi N, Parija SC, 2002. Uncultivable bacteria: implications and recent trends towards identification. Indian J. Med. Microbiol. 20, 174–177.

Chandra V, Torres M, Ortega YR, 2014. Efficacy of wash solutions in recovering *Cyclospora cayetanensis*, *Cryptosporidium parvum*, and *Toxoplasma gondii* from basil. J. Food Prot. 77, 1348–1354. [PubMed: 25198596]

Ekblom R, Wolf JB, 2014. A field guide to whole-genome sequencing, assembly and annotation. Evol. Appl. 7, 1026–1042. [PubMed: 25553065]

Gurevich A, Saveliev V, Vyahhi N, Tesler G, 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. [PubMed: 23422339]

Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, Ordoukhanian P, 2014. Library construction for next-generation sequencing: overviews and challenges. Biotechniques 56, 61–67. [PubMed: 24502796]

Kim H, Jebrail MJ, Sinha A, Bent ZW, Solberg OD, Williams KP, Langevin SA, Renzi RF, Van De Vreugde JL, Meagher RJ, Schoeniger JS, Lane TW, Branda SS, Bartsch MS, Patel KD, 2013. A microfluidic DNA library preparation platform for next-generation sequencing. PLoS One 8, e68988. [PubMed: 23894387]

Lalonde LF, Reyes J, Gajadhar AA, 2013. Application of a qPCR assay with melting curve analysis for detection and differentiation of protozoan oocysts in human fecal samples from Dominican Republic. Am.J.Trop. Med. Hyg. 89, 892–898. [PubMed: 24019437]

Lamble S, Batty E, Attar M, Buck D, Bowden R, Lunter G, Crook D, El-Fahmawi B, Piazza P, 2013. Improved workflows for high throughput library preparation using the transposome-based Nextera system. BMC Biotechnol. 13, 104. [PubMed: 24256843]

Lan JH, Yin Y, Reed EF, Moua K, Thomas K, Zhang Q, 2015. Impact of three Illumina library construction methods on GC bias and HLA genotype calling. Hum. Immunol. 76, 166–175. [PubMed: 25543015]

Langmead B, Salzberg SL, 2012. Fast gapped-read alignment with bowtie 2. Nat. Methods 9, 357–359. [PubMed: 22388286]

Lay MK, Atmar RL, Guix S, Bharadwaj U, He H, Neill FH, Sastry KJ, Yao Q, Estes MK, 2010. Norwalk virus does not replicate in human macrophages or dendritic cells derived from the peripheral blood of susceptible humans. Virology 406, 1–11. [PubMed: 20667573]

Morassutti AL, Perelygin A, De Carvalho MO, Lemos LN, Pinto PM, Frace M, Wilkins PP, Graeff-Teixeira C, DA Silva AJ, 2013. High throughput sequencing of the *Angiostrongylus cantonensis* genome: a parasite spreading worldwide. Parasitology 140, 1304–1309. [PubMed: 23863082]

Parkinson NJ, Maslau S, Ferneyhough B, Zhang G, Gregory L, Buck D, Ragoussis J, Ponting CP, Fischer MD, 2012. Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA. Genome Res. 22, 125–133. [PubMed: 22090378]

Qvarnstrom Y, Wei-Pridgeon Y, Li W, Nascimento FS, Bishop HS, Herwaldt BL, Moss DM, Nayak V, Srinivasamoorthy G, Sheth M, Arrowood MJ, 2015. Draft genome sequences from *Cyclospora cayetanensis* oocysts purified from a human stool sample. Genome Announc 19, e01324–15.

Rhodes J, Beale MA, Fisher MC, 2014. Illuminating choices for library prep: a comparison of library preparation methods for whole genome sequencing of *Cryptococcus neoformans* using Illumina HiSeq. PLoS One 9, e113501. [PubMed: 25409295]

van Dijk EL, Jaszczyszyn Y, Thermes C, 2014. Library preparation methods for next-generation sequencing: tone down the bias. Exp. Cell Res. 322, 12–20. [PubMed: 24440557]

Verweij JJ, Laeijendecker D, Brienen EA, van Lieshout L, Polderman AM, 2003. Detection of *Cyclospora cayetanensis* in travellers returning from the tropics and subtropics using microscopy and real-time PCR. Int J Med Microbiol 293, 199–202. [PubMed: 12868656]

Wang QP, Wu ZD, Wei J, Owen RL, Lun ZR, 2012. Human *Angiostrongylus cantonensis*: an update. Eur. J. Clin. Microbiol. Infect. Dis. 31, 389–395.

Yong HS, Eamsobhana P, Lim PE, Razali R, Aziz FA, Rosli NSM, Poole-Johnson J, Anwar A, 2015. Draft genome of neurotropic nematode parasite *Angiostrongylus cantonensis*, causative agent of human eosinophilic meningitis. Acta Trop. 148, 51–57. [PubMed: 25910624]

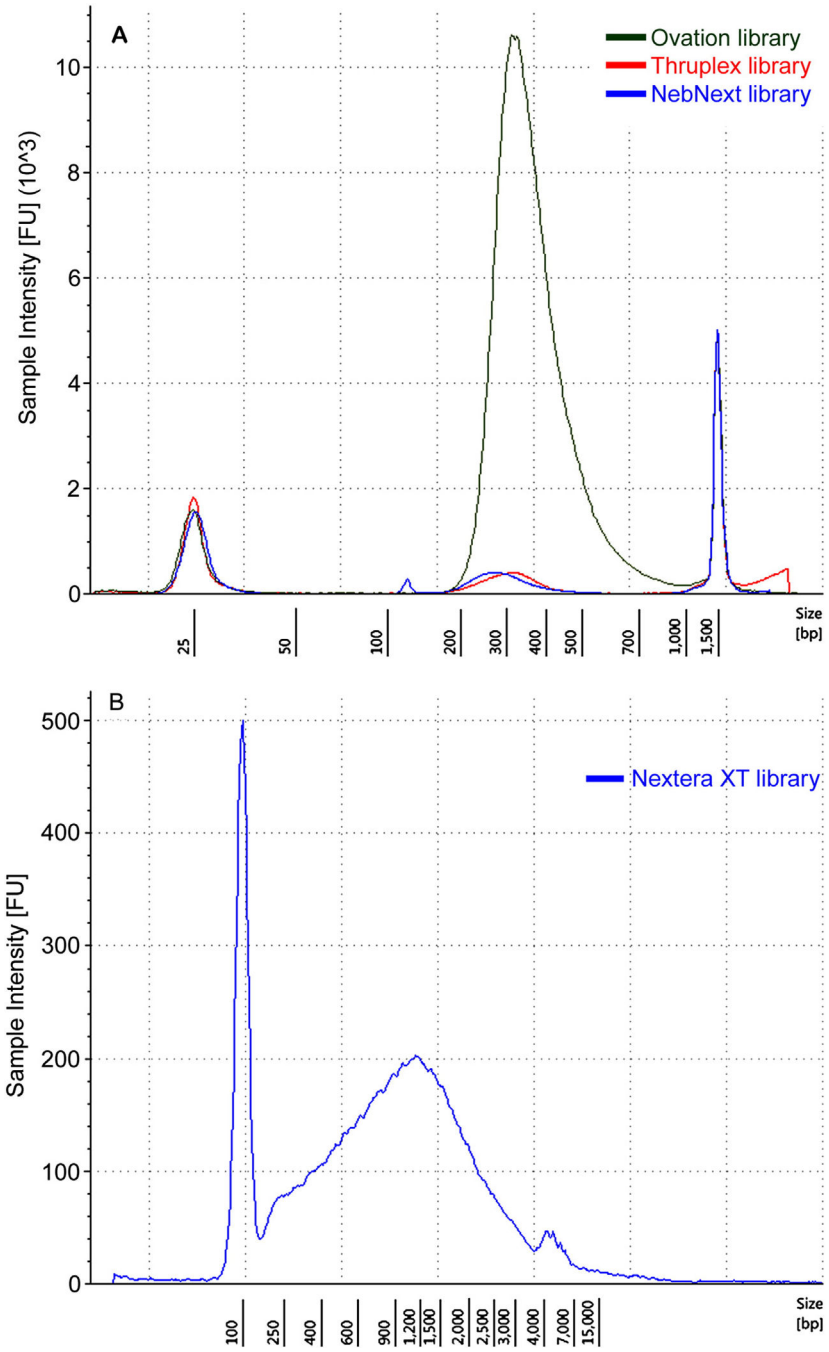

Fig. 1.

Size distribution and concentration of Illumina libraries of *Angiostrongylus cantonesis* using different methods. A): Ovation (green), NEBNext (blue) and ThruPlex (red) libraries as measured on a D1000 Screen Tape. The 25 and 1500 base pair peaks are internal size markers included in each lane. B): Nextera XT library as measured on a Genomic Screen Tape. The prominent 100 base pair peak is a size marker included in each lane.

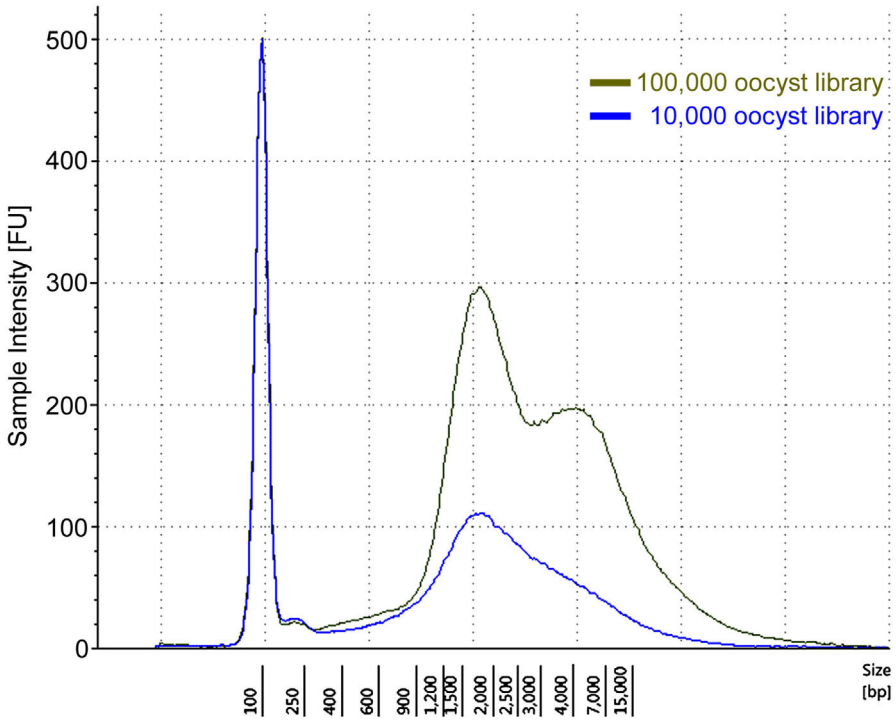

Fig. 2.

Size distribution and concentration of Ovation libraries of *Cyclospora cayetanensis* genomic DNA extracted from 100,000 (green) and 10,000 (blue) oocysts, measured on a Genomic Screen Tape. The prominent 100 base pair peak is a size marker included in each lane.

Table 1

Comparison of de *novo* assembly results of A. *cantonensis* sequences from different libraries.

Library	Geneious assembler					SPAdes assembler				
	Contigs	N50	Longest contig	Total length	CG%	Contigs	N50	Longest contig	Total length	CG%
NEBNext	502,648	272	13,496	147,851,316	42.4	193,124	404	21,077	80,929,040	42.1
Nextera	367,472	331	17,371	101,346,446	42.0	238,534	428	17,912	105,745,017	42.8
Ovation	260,382	407	17,396	87,134,446	41.2	159,728	530	41,900	89,039,577	41.7
ThruPlex	533,794	263	13,475	52,764,085	42.6	72,363	373	16,365	27,603,199	43.7