

1 **Title:** Rapid, High-Throughput, Cost Effective Whole Genome Sequencing of SARS-CoV-2  
2 Using a Condensed One Hour Library Preparation of the Illumina DNA Prep Kit

3  
4 **Authors:** Rebecca Hickman,<sup>1</sup> Jason Nguyen,<sup>1</sup> Tracy D. Lee,<sup>1</sup> John R. Tyson,<sup>1</sup> Robert Azana,<sup>1</sup>  
5 Frankie Tsang,<sup>1</sup> Linda Hoang,<sup>1,2</sup> Natalie Prystajeky<sup>1,2</sup>

6  
7 **Affiliations:**

8 <sup>1</sup>British Columbia Centre for Disease Control Public Health Laboratory, Vancouver, British  
9 Columbia, Canada

10 <sup>2</sup>The University of British Columbia, Vancouver, British Columbia, Canada

11  
12 **Running Title:** High-throughput, cost-effective SARS-CoV-2 sequencing

13 **Keywords:** SARS-CoV-2, Whole Genome Sequencing, Testing, Public Health, COVID-19

14  
15 **Corresponding Author:**

16 Natalie Prystajeky, PhD

17 BC Centre for Disease Control

18 655 West 12th Ave.

19 Vancouver, BC, Canada V5Z 4R4

20 natalie.prystajeky@bccdc.ca

21

## 22 **ABSTRACT**

23 The ongoing COVID-19 pandemic necessitates cost-effective, high-throughput, and timely  
24 genomic sequencing of SARS-CoV-2 viruses for outbreak investigations, identifying variants of  
25 concern (VoC), characterizing vaccine breakthrough infections, and public health surveillance.  
26 Additionally, the enormous demand of genomic sequencing on supply chains and the resulting  
27 shortages of laboratory supplies necessitate the use of low-reagent and low-consumable methods.  
28 Here, we report an optimized library preparation method where the same protocol can be used in  
29 a STAT scenario, from sample to sequencer in as little as eight hours, and a high-throughput  
30 scenario, where one technologist can perform 576 library preparations over the course of one 8-  
31 hour shift. This new method uses Freed et al.'s 1200 bp primer sets (Biol Methods Protoc  
32 5:bpaa014, 2020, <https://doi.org/10.1093/biomethods/bpaa014>) and a modified and truncated  
33 Illumina DNA Prep workflow (Illumina, CA, USA). Compared to the original, application of this  
34 new method in hundreds of clinical specimens demonstrated equivalent results to the full-length  
35 DNA Prep workflow at 45% the cost, 15% of consumables required (such as pipet tips), 25% of  
36 manual hands-on time, and 15% of on-instrument time if performing on a liquid handler, with no  
37 compromise in sequence quality. Results suggest that this new method is a rapid, simple, cost-  
38 effective, and high-quality SARS-CoV-2 whole genome sequencing protocol.

39

## 40 **INTRODUCTION**

41 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative pathogen for the  
42 novel 2019 coronavirus disease (COVID-19), is responsible for the pandemic first identified in  
43 Wuhan, China, and since has spread worldwide. As of June 2nd 2021, there have been over 171

44 million cases and 3.5 million COVID-19 deaths confirmed (<https://coronavirus.jhu.edu/>), and the  
45 pandemic continues to have devastating social, health, and economic impacts globally. The  
46 reliance on genomics for the study of viral transmission, identifying variants of interest (VoI)  
47 and of concern (VoC), and for studying vaccine breakthrough has necessitated rapid, high-  
48 throughput, and high-quality whole genome sequencing. This has resulted in unprecedented  
49 demand for laboratory reagents, consumables, equipment, and highly skilled laboratory staff to  
50 generate sequence data. With global shortages of laboratory reagents and consumables, there is  
51 widespread need for methods with low-reagent and low-consumable requirements that are  
52 scalable.

53 A multiplexed 1200 bp tiled primer amplicon approach previously described (1), which  
54 generates high and consistent coverage amplification across the SARS-CoV-2 genome, was  
55 chosen for best compatibility with the DNA Prep library preparation kit (Illumina, CA, USA)  
56 compared to other widely used primer sets (2, 3). Illumina's DNA Prep library preparation kit  
57 has been shown to generate high-quality, robust library preparations with a method that is  
58 scalable and liquid handler compatible. However, this library preparation method is  
59 comparatively expensive and more complex than alternatives. Here we use an optimized and  
60 truncated Illumina DNA Prep method that minimizes cost, hands-on time, and complexity of  
61 work while maintaining high-quality and robust sequence data.

62

## 63 **MATERIALS AND METHODS**

64 **SARS-CoV-2 RNA samples and real-time RT-PCR.** A total of 450 clinical samples tested  
65 positive for SARS-CoV-2 by various validated laboratory methods throughout the province of

66 British Columbia (BC) were forwarded to the BC Centre for Disease Control Public Health  
67 Laboratory (BCCDC PHL) for whole genome sequencing. These samples—collected by either a  
68 nasopharyngeal (NP) swab in Yocon transport media or 0.9% saline gargle—were extracted at  
69 the BCCDC PHL using ThermoFisher’s Viral RNA extraction kit on the  
70 MagMAX™/KingFisher™ Flex extractors.

71 **cDNA synthesis and tiled amplicon generation.** cDNA synthesis and amplicon generation  
72 were performed as previously described by Freed et al. in steps 1-13 on the protocols.io web-  
73 based platform ([dx.doi.org/10.17504/protocols.io.bgc8jszw](https://doi.org/10.17504/protocols.io.bgc8jszw)).

74 **Library preparation and Illumina sequencing.** In the originally validated protocol, working in  
75 sets of 96 samples, the two multiplex primer pools were combined and purified at a 1:1 volume  
76 with AMPure XP beads (Beckman Coulter, IN, USA) and two sequential washes with 80%  
77 ethanol. The library preparation was performed with the purified amplicons as per the  
78 manufacturer’s recommendations (4), which consists of tagmentation, tagmentation stop, three  
79 washes, a reduced 5-cycle library PCR with indexing, and a dual-sided size selection and  
80 purification, before individual libraries are pooled (Fig 1A).

81 The optimized library preparation method, performed at half reaction volumes, eliminates  
82 amplicon purification, tagmentation stop, and post-tagmentation washes entirely (Fig 1B).  
83 Additionally, pooling was performed on the indexed post-PCR products, rather than the libraries,  
84 allowing the dual-sided size selection and purification steps to be performed per 96-sample pool  
85 instead of per sample.

86 Briefly, the multiplex primer pools were combined. Next, 15µl of amplicon, 5µl of bead linked  
87 transposomes (BLT) and 5µl of tagmentation buffer 1 (TB1) were heated on a thermal cycler at

88 55°C for 15 minutes. On a magnet, the supernatant was removed and taken off the magnet, the  
89 beads were then re-suspended in 10µl of enhanced PCR mix (EPM), 10µl of ultrapure water, and  
90 5µl of Illumina’s Unique Dual Indexes before being run on a reduced 8-cycle library PCR. After  
91 PCR, equal volumes of each library supernatant were pooled. In a single microcentrifuge tube,  
92 45µl of the post-PCR pool was added to 40µl of ultrapure water and 45µl sample purification  
93 beads (SPB). After five minutes, 125µl of the supernatant was transferred to a new  
94 microcentrifuge tube with 15µl of SPB. After a final 5-minute incubation, two sequential 80%  
95 ethanol washes were performed on the magnetic beads before the tagmented, size selected, and  
96 purified library was eluted in resuspension buffer (RSB). The finished library was quantified by  
97 Qubit, denatured, and diluted to a final concentration of 15pM, and 96 samples were sequenced  
98 using the Illumina MiSeq system using a 2 x 150 bp MiSeq Reagent Kit v2 Micro. A detailed  
99 laboratory method can be found on protocols.io ([dx.doi.org/10.17504/protocols.io.b3vgqn3w](https://doi.org/10.17504/protocols.io.b3vgqn3w)).

100 **epMotion Automation.** The five runs for this report were performed manually; however, this  
101 protocol has since been automated on the Eppendorf epMotion 5075t liquid handler and has been  
102 used to prepare over 50,000 libraries. Using only four boxes of pipet tips, this method is on the  
103 instrument for 45 minutes with user intervention only required to transfer the plate to a thermal  
104 cycler. The epMotion protocol is available on request.

105 **Data analysis.** The data was analyzed using a modified Nextflow bioinformatics pipeline  
106 (<https://github.com/BCCDC-PHL/ncov2019-artic-nf>) from the Simpson Lab  
107 (<https://github.com/jts/ncov2019-artic-nf>) that is built upon the original Connor laboratory’s  
108 ncov2019-artic-nf Nextflow pipeline (<https://github.com/BCCDC-PHL/ncov2019-artic-nf>).

109 Reports describing sequencing quality metrics—including genome completeness, depth of  
110 coverage, and quality flags—and lineage information were generated using ncov-tools from the

111 Simpson Lab (<https://github.com/jts/ncov-tools>). Lineages were assigned according to the  
112 Pangolin version 2.4.2 (5). Resource utilization, estimated in terms of time, cost, and  
113 consumables used, was compared between the new and original sequencing methods to  
114 determine potential savings.

115

## 116 RESULTS

117 **Illumina quality metrics.** The basic quality metrics used for Illumina sequencing remained high  
118 and met the expected values on sequencing runs using the new method. Over five MiSeq runs,  
119 the original method had a mean 94.2% of reads with a Q-score >30 (range 92.7 to 95.0%),  
120 compared to the new method which had a mean 95.1% of reads with a Q-score >30 (range 94.2  
121 to 95.8%).

122 **Sequencing depth of coverage and genome completeness.** The mean sequencing depth on the  
123 full-length DNA Prep method and the new reduced library preparation method were 412X  
124 coverage and 433X coverage, respectively. Using the original method, 387/450 (86.0%) of  
125 sequences passed quality metrics, where a pass is >85% of the genome sequenced to >10X depth  
126 (Table 1). In comparison, 389/450 (86.4%) sequences passed the quality metrics by the new  
127 method. By the original method, 363/450 (80.7%) sequences had over 99% of the SARS-CoV-2  
128 genome and by the new method, 366/450 (81.3%) sequences had over 99% genome coverage  
129 (Table 1). Of the 67 samples that did not pass the sequencing metrics on either method, 43  
130 (64.2%) had Ct values available, of which 43/43 (100%) had a Ct value >28 and 41/43 (95.3%)  
131 had a Ct value >30. The mean Ct value of a sample that did not pass the quality metrics was 34.3  
132 (Fig 2).

133 **Lineage calls.** A >70% genome completeness was required to identify the SARS-CoV-2 lineage.  
134 In total, 391/450 (86.8%) of libraries from the original method and 392/450 (87.1%) of libraries  
135 from the new method had sufficient genome completeness to assign a lineage (Table 1). In the  
136 386 samples that had a lineage assignment by both methods as well as a passing >85% coverage  
137 and no quality flags, 383 (99.2%) of lineages matched between methods. The three samples with  
138 lineage mismatches had 3-13% of their genomes unsequenced, with the missing locations being  
139 key mutations sites that altered the lineage calls.

140 **Resource utilization.** The new method required less total time to complete compared to the  
141 original method, both when performed manually (2 hours vs. 6 hours, respectively) and on an  
142 Eppendorf epMotion 5075t instrument (1.5 hours vs. 6.5 hours, respectively) (Table 2, Fig 1B).  
143 The new method was also associated with a lower cost of library preparation (\$31 vs. \$68) and  
144 required fewer tip boxes (4 vs. 34) compared to the original method (Table 2, Fig 1B).

145

## 146 **DISCUSSION**

147 Sequencing has played an invaluable role in the response to the COVID-19 pandemic.  
148 Specifically, sequencing has been used to complement outbreak investigations, to study viral  
149 transmission and introductions, to identify emerging variants, and to understand vaccine  
150 breakthrough infection. Ongoing work in this area, however, demands optimization of laboratory  
151 work for whole genome sequencing in order to increase sequencing capacity, improve turn-  
152 around-time, and reduce cost without compromising sequence quality. This report describes a  
153 new sequencing method that incorporates workflow modifications that save time, limit the use of  
154 reagents and consumables, and maintain high-quality sequencing results.

155 **Time savings.** The original Illumina DNA Prep library preparation, performed manually,  
156 requires almost a full 8-hour work shift to complete, the vast majority of it being hands-on time.  
157 In addition, the multiple wash steps are difficult on the laboratory worker and prone to error. The  
158 original method was programmed on an Eppendorf epMotion 5075t, but took 6.5 hours on the  
159 instrument to complete. Both of these options were slow and laborious for the laboratory workers  
160 and instruments. The new method removes amplicon purification, tagmentation washes, and  
161 pools the dual sided size selection into a single tube thereby removing the vast majority of the  
162 pipetting steps. A single laboratory worker is now capable of producing 576 library preparations  
163 by this method manually in an 8-hour shift. Furthermore, when performed on the epMotion  
164 instrument, the entire method takes only 45 minutes for a 96-well plate, freeing the instrument  
165 for the next library preparation and allowing one instrument to perform 7+ 96-well plate libraries  
166 in an 8-hour shift. With the same instruments and staff, we were able to sequence 6-8X as many  
167 libraries as the previous method in the same amount of time. This is particularly important during  
168 staff shortages or during testing surges when human resources are a limiting factor.

169 In addition, as the library preparation takes one hour, it allows for the unique opportunity of  
170 going from sample to sequencer in the course of one shift in a STAT scenario. Extraction, cDNA  
171 synthesis, and amplicon generation typically take around six hours; however, our reduced library  
172 preparation method allows the samples to be on the sequencer by the end of the shift.

173 **Reagent and consumable savings.** During a time of global laboratory reagent and consumable  
174 shortages, the elimination of multiple steps from the original protocol has other positive effects.  
175 Removing the amplicon clean up and halving the reaction volume of the DNA Prep library  
176 preparation reduced the reagent cost of whole genome sequencing by nearly 55%, making an  
177 initially expensive library preparation kit more affordable. Additionally, removing the majority



178 of steps, especially washes, reduces pipet tip usage substantially from 34 tip boxes to four (an  
179 85% reduction) for a 96-sample library preparation.

180 **High quality results.** Despite the removal of the amplicon purification and the majority of  
181 library preparation steps, results have shown that the quality of the sequencing was not  
182 compromised. The percent of samples with complete genomes (>99%) and passing quality  
183 control cutoffs (>85% genome completeness) were equivalent between the two methods. Of  
184 samples passing quality control, 99.2% of the 450 samples had a matching lineage between the  
185 methods; the three that did not have matching lineages had 3-13% of the genome missing in key  
186 lineage-defining sites.

187 In conclusion, this study demonstrates a rapid library preparation method for SARS-CoV-2  
188 whole genome sequencing that produces consistent and high-quality data with equivalent results  
189 to the original full length DNA Prep library preparation as described from the manufacturer.  
190 With an overall savings of >75% hands-on time at 45% the cost and using 15% of the  
191 consumables compared to a typical DNA Prep library preparation, this new protocol is an  
192 efficient, scalable, and pragmatic alternative for SARS-CoV-2 whole genome sequencing.

193

194

195

196

197

198

199 **AUTHORS' CONTRIBUTIONS**

200 RH designed the experiment, tested the protocol, analyzed the data, and wrote the manuscript. JN  
201 conceived the idea for the high throughput method, assisted with the development, testing, and  
202 automation of the protocol, and helped with analyzing the initial developmental data. TDL and  
203 JT contributed to experimental design. FT and RA provided operational support and coordination  
204 for implementation of the protocol. NP and LH provided study oversight. NP, JT, and TDL  
205 critically reviewed the manuscript and all authors approved the final draft.

206 **ACKNOWLEDGEMENTS**

207 We would like to thank the Virology Laboratory at the BCCDC PHL for performing routine  
208 diagnostic testing of SARS-CoV-2. We would like to thank the Bacteriology and Mycology  
209 Laboratory at the BCCDC PHL for performing the routine whole genome sequencing of SARS-  
210 CoV-2. We would also like to the BCCDC data analytics and bioinformatics team for pipeline  
211 design and optimization, and Jessica Caleta for assistance with figure design.

212 **ETHICS STATEMENT**

213 This study was conducted under the auspices of a quality improvement initiative, authorized by  
214 the Provincial Health Officer under the Public Health Act. Ethical oversight was waived by the  
215 University of British Columbia Clinical Research Ethics Board, in accordance with Article 2.5 of  
216 the Tri-Council Policy Statement 2 (TCPS 2).

217 **CONFLICTS OF INTEREST**

218 The authors declare no conflicts of interest.

219 **FUNDING**

220 This study did not receive any funding.

221 **REFERENCES**

- 222 1. Freed NE, Vlková M, Faisal MB, Silander OK. 2020. Rapid and inexpensive whole-  
223 genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford  
224 Nanopore Rapid Barcoding. *Biol Methods Protoc* 5:bpaa014.
- 225 2. Quick J. 2020. nCoV-2019 sequencing protocol v3 (LoCost).  
226 <https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye>
- 227 3. Tyson JR, James P, Stoddart D, Sparks N, Wickenhagen A, Hall G, Choi JH, Lapointe H,  
228 Kamelian K, Smith AD, Prystajeky N, Goodfellow I, Wilson SJ, Harrigan R, Snutch TP,  
229 Loman NJ, Quick J. 2020. Improvements to the ARTIC multiplex PCR method for  
230 SARS-CoV-2 genome sequencing using nanopore. *bioRxiv*:2020.09.04.283077.
- 231 4. Illumina. 2020. Nextera DNA Flex Library Preparation Reference Guide, document  
232 1000000025416 v07. Illumina, San Diego, CA.  
233 <https://sapac.support.illumina.com/content/dam/illumina->  
234 [support/documents/documentation/chemistry\\_documentation/illumina\\_prep/illumina-](https://sapac.support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/illumina_prep/illumina-)  
235 [dna-prep-reference-guide-1000000025416-09.pdf](https://sapac.support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/illumina_prep/illumina-dna-prep-reference-guide-1000000025416-09.pdf)
- 236 5. Rambaut A, Holmes EC, O’Toole Á, Hill V, McCrone JT, Ruis C, du Pleiss L, Pybus  
237 OG. 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist  
238 genomic epidemiology. *Nat Microbiol* 5:1403–1407.

239

240

241

242

243 **FIGURE LEGENDS**

244 **Fig 1.** Comparison of step-wise workflow, total time to complete, and tip box usage between the  
245 (A) original whole genome sequencing method and the (B) new method  
246 ([dx.doi.org/10.17504/protocols.io.b3vgqn3w](https://doi.org/10.17504/protocols.io.b3vgqn3w)).

247 **Fig 2.** Percent genome completeness for the new method (triangles) and original method (circles)  
248 based on available cycle threshold (Ct) values.

249

250

251

252

253

254

255

256

257

258

259

260

261

**Table 1.** Genome completeness for each library method

	<b>Original method</b>	<b>New method</b>
Number of samples	450	450
% Passed quality metric (>85%)	86.0 (387/450)	86.4 (389/450)
% Complete genome (>99%)	80.7 (363/450)	81.3 (366/450)
% Lineages called (>70%)	86.8 (391/450)	87.1 (392/450)

262

263

**Table 2.** Time, cost, and consumable savings

	<b>Original method</b>	<b>New method</b>
Hands-on time (manual)	4 hours, 30 minutes	1 hour
Total time to completion (manual)	6 hours	2 hours
Time on instrument (epMotion)	6 hours	45 minutes
Total time to completion (epMotion)	6 hours, 30 minutes	1 hour, 30 minutes
Cost of library prep (CAD list price)	\$68	\$31
Number of tip boxes	34	4

264

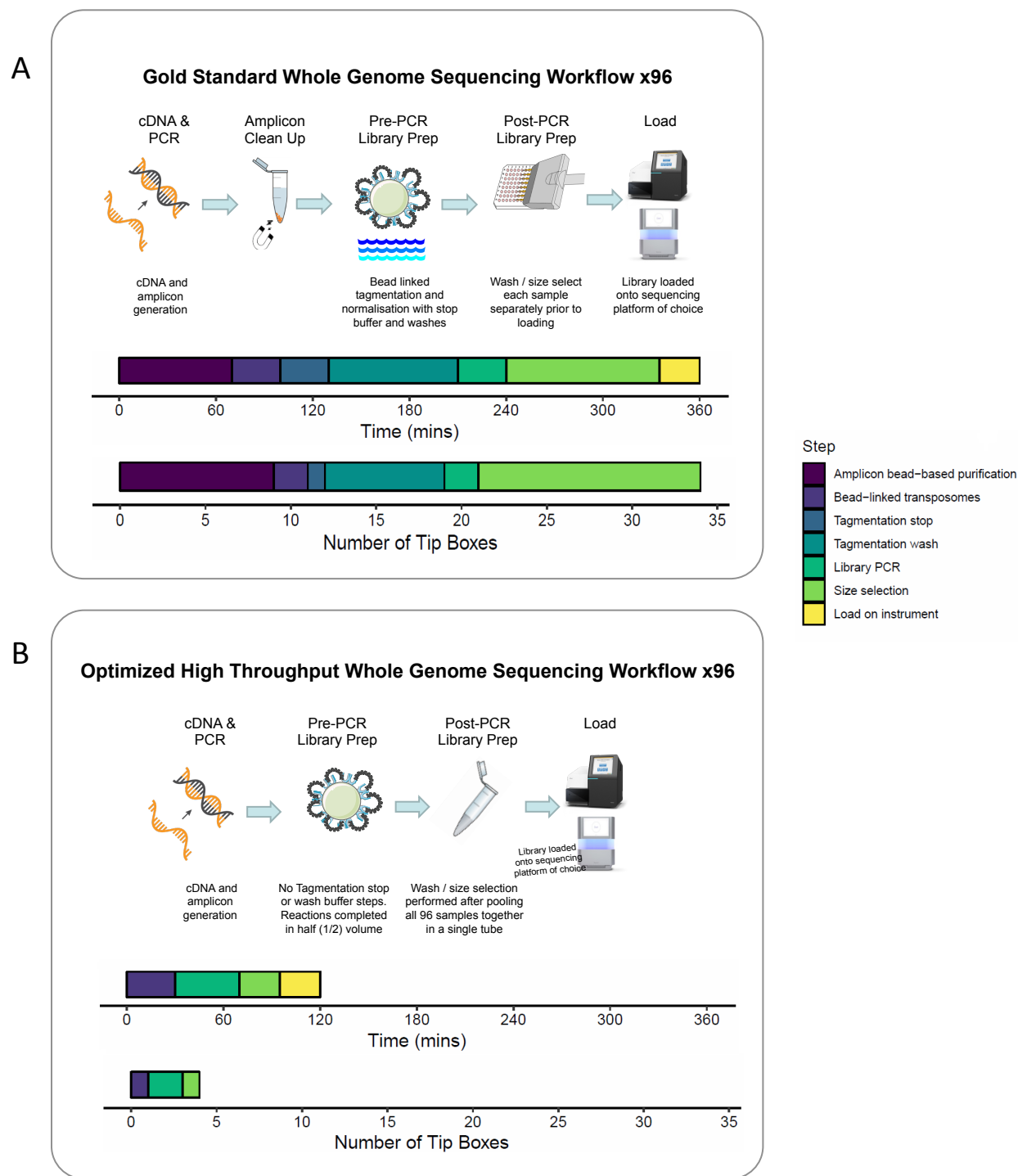
265

266

267

268

269

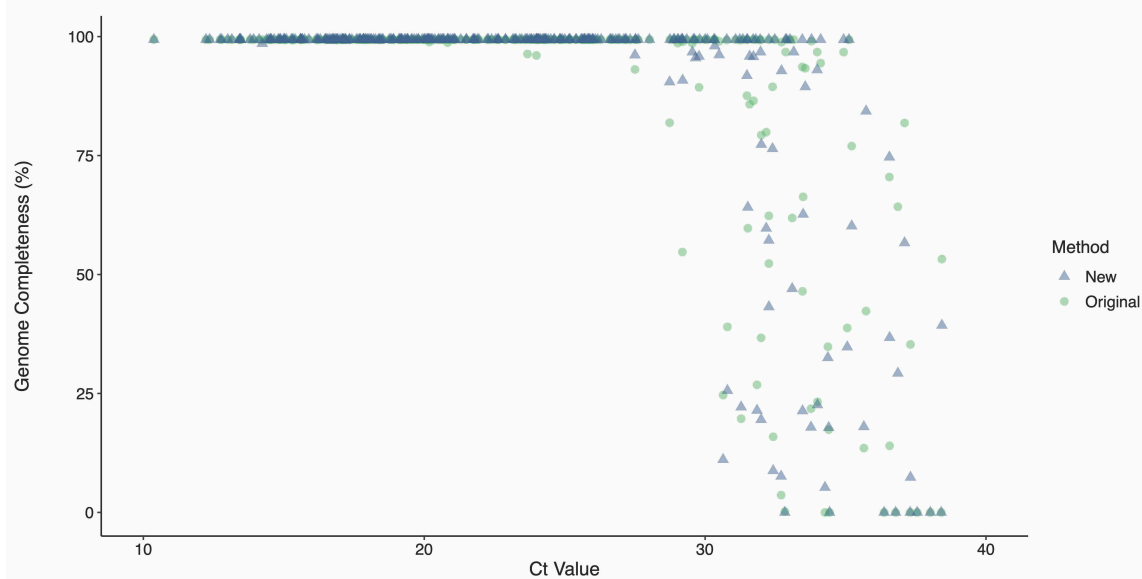


270

271 **Fig 1.** Comparison of step-wise workflow, total time to complete, and tip box usage between the

272 (A) original whole genome sequencing method and the (B) new method

273 ([dx.doi.org/10.17504/protocols.io.b3vgqn3w](https://doi.org/10.17504/protocols.io.b3vgqn3w)).



274

275 **Fig 2.** Percent genome completeness for the new method (triangles) and original method (circles)

276 based on available cycle threshold (Ct) values.