GINKGO
AUTOMATION

High-throughput, Miniaturized Next-generation Sequencing (NGS) Library Preparation on the Ginkgo Reconfigurable Automation Cart (RAC) Platform

---

30-40%

Decrease in hands-on-time

33%

Reduction in required headcount

10-fold

Savings on overall reagent costs

WWW.AUTOMATION.GINKGO.BIO

High-throughput, Miniaturized Next-generation Sequencing (NGS) Library Preparation on the Ginkgo Reconfigurable Automation Cart (RAC) Platform

Paulina Kanigowska, Ph.D., Elizabeth Gendreau, Teryn Citino, Alicia Byrn, Brintha Girinathan, Aimable Habimana, Anna Peters, Noah Nsangou, Alycia Wong

Here, we show how we successfully miniaturized (100-fold) and adapted the Illumina DNA Prep workflow to be end-to-end executed on Ginkgo's RAC platform, including upstream microbial cell lysis, thereby enabling quick, crude, and low-cost (~$3/sample) DNA sequence verification of 100s to 1000s purified and unpurified input samples daily, with minimal human intervention.

INTRODUCTION

High-throughput genetic engineering necessitates frequent cost- and time-efficient iteration through the Design, Build, Test and Learn (DBTL) cycles. In recent years, next-generation sequencing (NGS) has significantly improved this iteration, becoming the go-to Test approach for verifying both successful DNA assembly of plasmids and genome engineering of microbial strains, at both academic and industrial biotech facilities[2,3].

Previously published studies have been primarily focusing on automation-enabled miniaturization of a more automation-friendly, but less flexible (with respect to the input sample amount), Nextera® XT DNA Library Prep workflow[2,3]. The most recent studies attempted to miniaturize the Illumina DNA Prep workflow 10-fold, albeit manually[1]. Our work demonstrates how careful biological workflow adaptation towards a fully automated solution and a robust automation platform can improve upon these efforts.

FULLY AUTOMATED BIOLOGICAL WORKFLOW

The fully automated biological workflow consisted of the following main steps (see **FIG. 1**), previously executed in a partially automated manner (using two benchtop liquid handling workstations - Beckman Coulter Echo 525 and Hamilton Microlab STAR):

1. **Yeast cell wall lysis**: 37℃ 1 hr 450 rpm shaking incubation with 155 µL of yeast cell

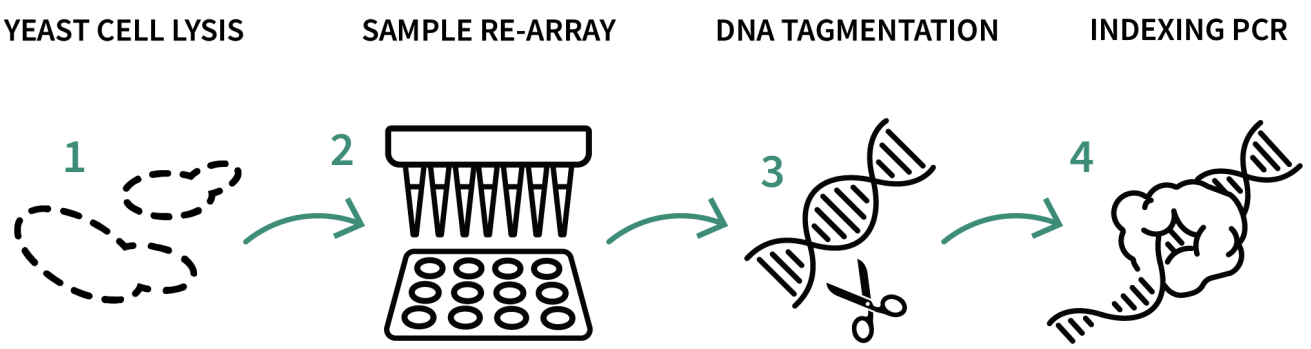


FIGURE 1. **Fully Automated NGS Library Prep Workflow**

lysis mix A (set up with Agilent Multiflo FX and Inheco Single Plate Incubators)

2. **Protein digestion:** 50℃ 30 min incubation with 40 µL yeast cell lysis mix B and 70℃ 1 hr yeast cell lysis mix B inactivating incubation (set up with Agilent Multiflo FX and Inheco Single Plate Incubators)
3. **96-well to 384-well microplate sample re-array:** Sample re-array from eight (8) 96-well microplates into two (2) Echo Qualified 384-well Polypropylene Microplates (set up with Agilent Bravo)
4. **DNA tagmentation reaction setup & incubation:** Nanoliter dispensing 300 nL of sample (crude yeast cell lysate, purified plasmid DNA, or nuclease-free water no template control), 100 nL of buffer-exchanged Bead-linked Transposomes (BLTs) and 100 nL of TB1 buffer into each well of 2 384-well PCR microplates, and 55℃ 15 min incubation (set up with Beckman Coulter Echo 525 and Thermo Fisher Scientific Automated Thermal Cyclers)
5. **Indexing PCR reaction setup & thermal cycling:** Nanoliter dispensing 125 nL of SDS (for quenching the DNA tagmentation), followed by nanoliter dispensing 125 nL of fwd and 125 nL of rev CDI indexing primers, bulk dispensing 12 µL of PCR master mix and thermal cycling - 1 cycle of 72℃ for 3 min and 98℃ for 30 sec, 15 cycles of 98℃ for 10 sec, 63℃ for 30 sec and 72℃ for 30 sec, 1 cycle of 72℃ for 2 min and 4℃ for 30 sec (set up with Beckman Coulter Echo 525, Agilent Multiflo FX and Thermo Fisher Scientific Automated Thermal Cyclers)

FIG. 2 shows Ginkgo's Automation Control Software (ACS) biological workflow scheduling results. The biological workflow was divided into three independently re-usable ACS protocols, which were altogether **executed in ~5 hrs 45 min, processing 768 samples without any in-person monitoring and runtime issues. Standard preventative online monitoring was performed by Ginkgo's Managed Automation Solution (MAS) team**.

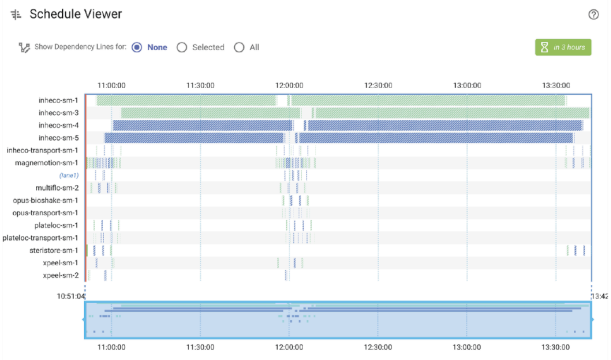
As expected, the Yeast Cell Lysis ACS protocol runs constituted most (~3 hrs) of the total turnaround time, due to multiple 0.5-1 hrs-long incubation steps. The NGS Library Prep ACS protocol runs were the second longest (~2 hrs), and required tightly controlled, timely execution of nanoliter scale DNA tagmentation steps, which was ultimately ensured via the ACS protocol time constraints feature. **Overall, full end-to-end automation reduced the number of operators running the biological workflow daily from 3 to 2, with the frequent manual loading / unloading of the Echo 525 acoustic nanoliter dispenser being one of the team's most laborious lab operations.**

NGS LIBRARY QUALITY CONTROL

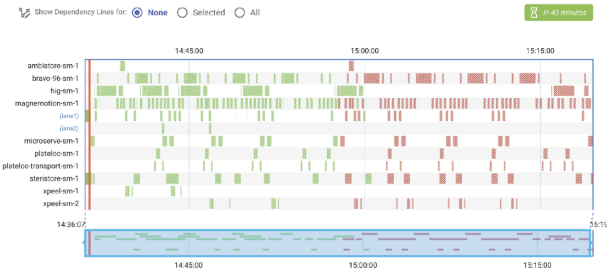
To achieve fast and low per-sample cost NGS for plasmid DNA and microbial genomes at scale, we intentionally balanced increasing throughput and miniaturization with read depth variability. Flow cell capacity was configured to reliably achieve a minimum of 30x average coverage by targeting an average of 100x read depth coverage. This approach allows us to accommodate unpurified, variable concentration input samples, such as yeast cell lysates, without requiring individual, unpooled NGS libraries purification or normalization.

A subset of newly constructed, unpurified, unnormalized, and unpooled NGS libraries was thus first analyzed via capillary electrophoresis (set up with Agilent TapeStation) - eight (8) NGS

1. YEAST CELL LYSIS ACS PROTOCOL RUNS



2. SAMPLE RE-ARRAY ACS PROTOCOL RUNS



3. NGS LIBRARY PREP ACS PROTOCOL RUNS (DNA TAGMENTATION & INDEX PCR)

FIGURE 2. Yeast Cell Lysis, Sample Re-array and NGS Library Prep Automation Control Software (ACS) Protocol Runs; Two (2) Yeast Cell Lysis ACS protocol runs were executed in ~3 hrs, each processing two (2) 96-well microplates; two (2) Sample Re-array ACS protocol runs were executed in ~45 min, each processing one (1) 384-well microplate; two (2) NGS Library Prep ACS protocol runs were executed in ~2 hrs, each processing one (1) 384-well microplate

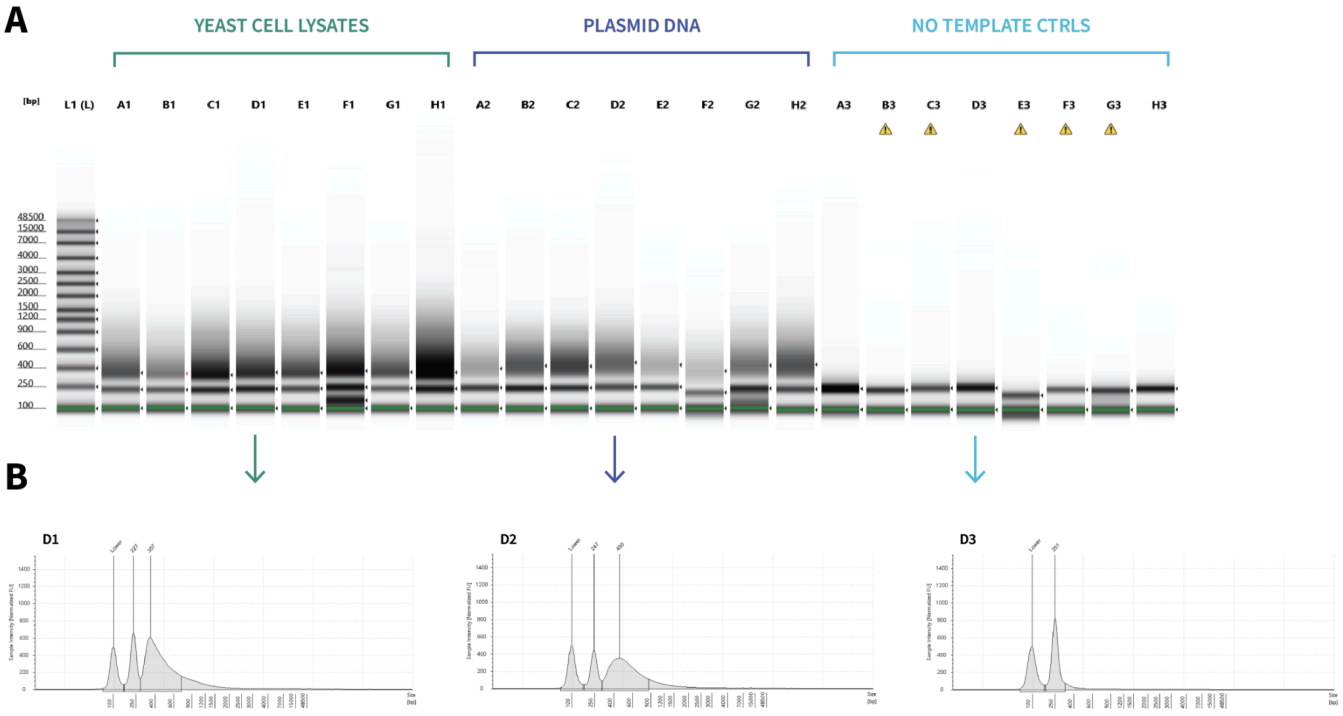


FIGURE 3. Preliminary QC of a Subset of 3 Different Input Sample Type Unpooled and Unpurified NGS Libraries; **A** Gel images of 3 different input sample type NGS library sets (8 NGS libraries / each sample type); **B** Electropherograms of example NGS libraries - one / each input sample type set (yeast cell lysate NGS library - target DNA fragment size distribution: mean DNA fragment size = 357 bp, DNA concentration = 48.1 ng / µL; plasmid DNA NGS library - target DNA fragment size distribution: mean DNA fragment size = 450 bp, DNA concentration = 32.7 ng / µL; no template control NGS library - only the off-target DNA fragment size distribution present)

libraries per each input sample type (yeast cell lysate, plasmid DNA and no template control). This preliminary analysis revealed expected target NGS library DNA fragment size distributions, with a mean DNA fragment size between ~300 and ~400 bp. All NGS libraries harbored a second distinct, and also expected (pre-purification), contaminating DNA fragment population (mean DNA fragment size of ~200 bp), likely corresponding to stable indexing PCR primer secondary structures (see **FIG. 3**).

Following the preliminary analysis, equal NGS library volumes were pooled into the three input sample type pools and subjected to double-sided size-selection via magnetic bead purification, which successfully enriched the target NGS libraries, as seen on the capillary electrophoresis electropherograms (see **FIG. 4**).

The pools were subsequently normalized to the same DNA concentration, pooled at a ratio ensuring at least 30x average coverage of each NGS library, considering different input sample

DNA sizes (*S. cerevisiae* CEN.PK genome size = 12.07 Mb, pUC19 plasmid DNA size = 2.68 kb), and then spiked-in with a PhiX positive control, denatured and neutralized (following Illumina's recommendations).

NGS RUN QUALITY CONTROL & RESULTS

Illumina's NovaSeq 6000 S2 reagent kit and flow cell were used to DNA sequence-verify the obtained NGS libraries (2 × 150 cycles paired-end DNA sequencing). The NGS run passed Illumina's QC criteria (yield = 0.39 Tbp, %Q30 = 84.32, %PF = 64.10). The vast majority of plasmid DNA and yeast cell lysate input samples generated sufficient read counts to achieve 30x average depth of coverage, with only 2% of yeast cell lysate samples generating fewer reads. All plasmid DNA samples achieved the target 30x average depth of coverage. For DNA sequencing applications at this scale, unique dual indexing can be cost prohibitive, so combinatorial dual indexing was used insead, which does allow for a low level of index hopping on Illumina flow cells. Regardless, we observed no false positive results and marginal mean read counts being generated for our no template control samples. Mapped (against the pUC19 plasmid DNA sequence) mean read counts were 3-4 orders of magnitude lower as compared to the pUC19 plasmid DNA and yeast cell lysate samples, respectively (see **FIG. 5-7**). As expected, given our previously described quick, crude and low-cost DNA sequencing approach, the total read count %CVs for plasmid DNA and yeast cell lysate input sample NGS libraries were elevated and ranged between 30-50%.

Interestingly, we found that the additional flow cell capacity expense, compensating for

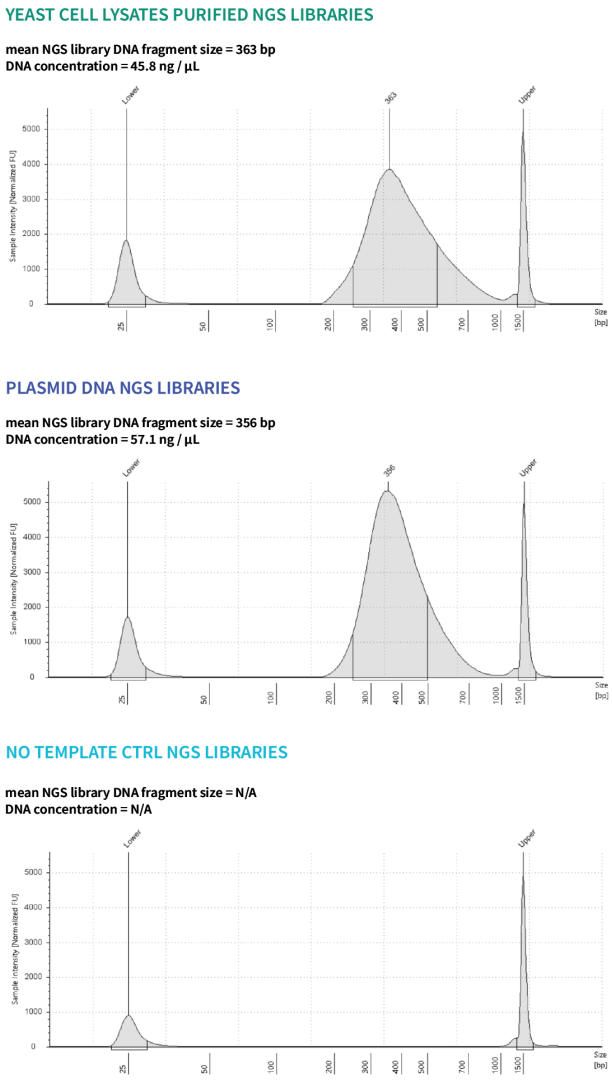


FIGURE 4. QC of 3 Different Input Sample Type's Pooled and Purified NGS Libraries; Plasmid DNA and yeast cell lysate input sample NGS library mean DNA fragment sizes and DNA concentration are shown; No valid NGS library was generated for no template control input samples

increased process variability, was overall lower as compared to the baseline flow cell capacity cost, coupled with the necessarily more stringent input and output sample pre- and post-processing.

ADAPTING BIOLOGICAL WORKFLOWS TOWARDS FULL AUTOMATION

While the Illumina DNA Prep workflow provides significant flexibility with respect to input DNA sample type, purity and quantity, its magnetic bead-linked transposome (BLT) biochemistry poses certain challenges to full biological workflow automation. Therefore, most studies to date have been focusing on automating and optimizing the less flexible, but more automation-friendly, Illumina Nextera® XT workflow, relying on in-solution transposome biochemistry[2,3].

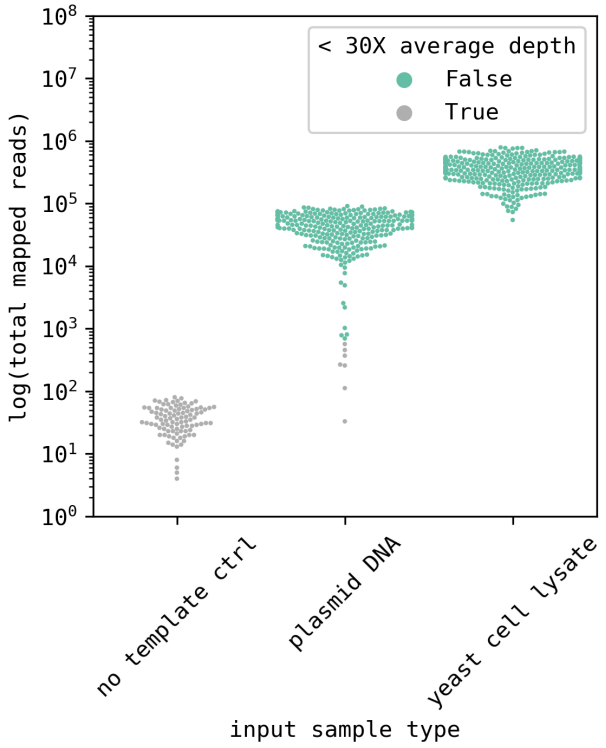


FIGURE 5. Swarmplot of Total Mapped Reads Distributions for Each of the 3 Input Sample Types; Samples with sufficient mapped reads to achieve ≥30x average depth of coverage are highlighted in green; Plasmid DNA and yeast cell lysate input sample total mapped reads %CVs are 49.79% and 39.65%, respectively

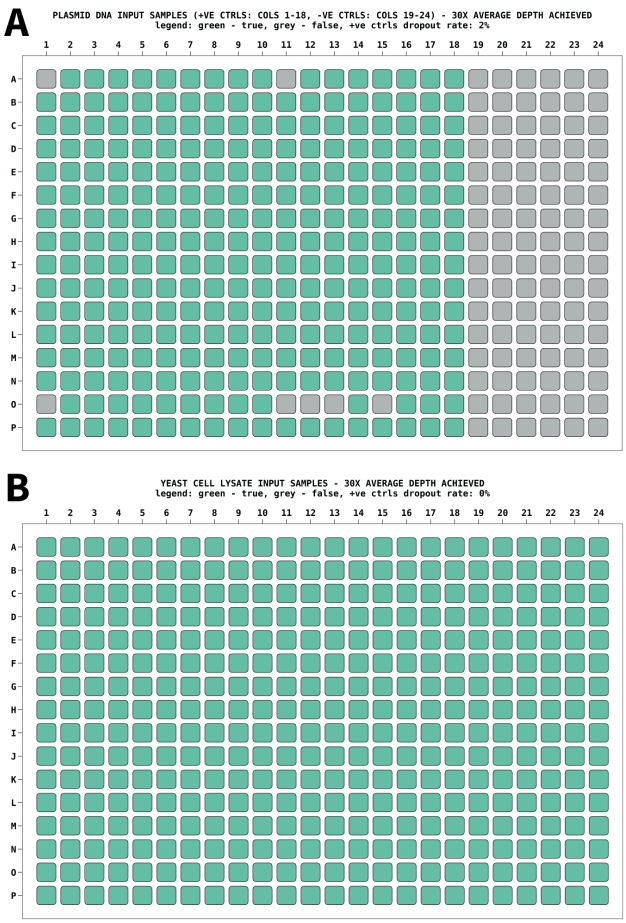


FIGURE 6. Microplate Maps of Success / Failure in Achieving the Target 30x Average Depth of Coverage; Dropouts (failures to meet at least 30x average depth of coverage) are highlighted in gray - the dropout rate for **A** plasmid DNA and **B** yeast cell lysate input sample NGS libraries is 2% and 0%, respectively; none of our no template controls showed false positive results (see microplate columns 19-24 in **A**; The results shown here are overall comparable to the previous partially manual and partially automated workflow execution

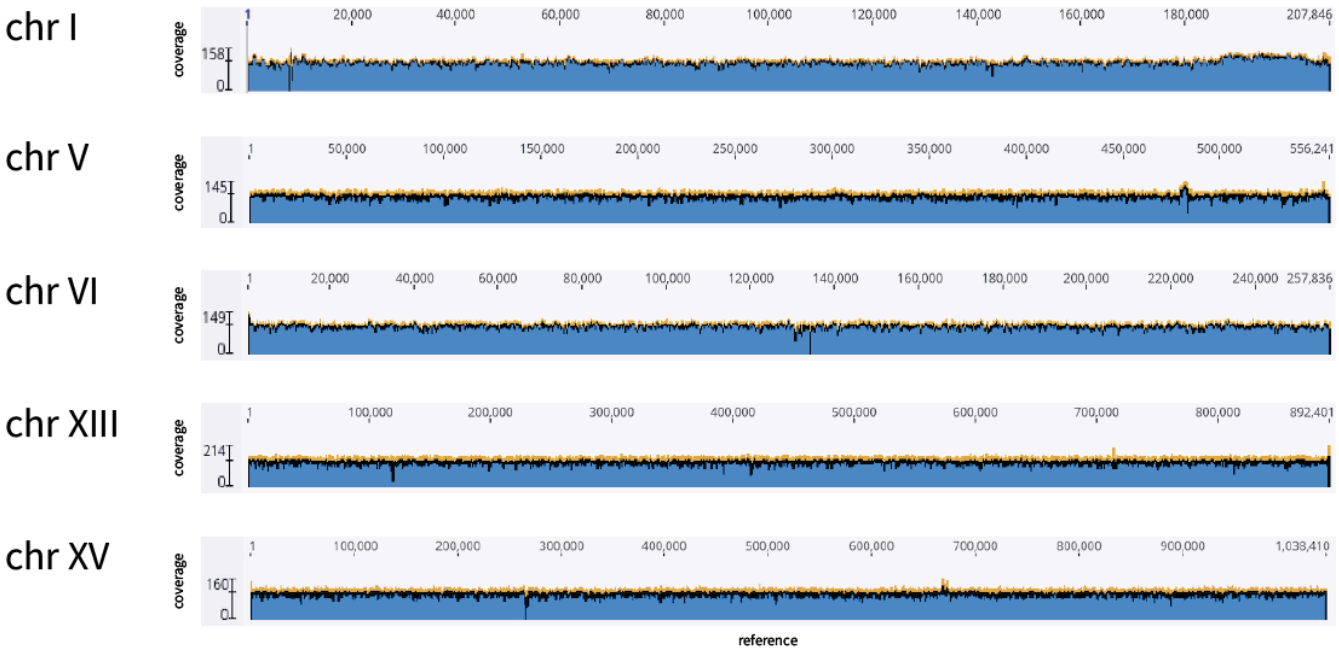


FIGURE 7. A Snapshot of Coverage Uniformity for a Representative Yeast Cell Lysate Sample and a Subset of Chromosomes; 5 out of 12 *S. cerevisiae* (CEN.PK strain) chromosome reference alignments (Geneious software screenshots shown)

Additionally, Illumina's recommended reaction / processing volumes translate to high (~$30 per input sample) costs, which are justifiable for low-throughput laboratory operations, but became prohibitive for our microbial genome and plasmid DNA sequence verification purposes.

We first focused on reaction volume miniaturization. On the RAC platform, we achieved it by using an integrated Beckman Coulter's Echo 525 acoustic nanoliter dispenser, which ultimately enabled us to reduce the reagent costs 10-fold, by miniaturizing key reaction volumes (most notably the tagmentation reaction volume), without sacrificing the liquid handling accuracy.

Accommodating increasingly large, and thus longer duration, NGS library prep workflows was however challenging, with the BLT sedimentation being one of the main issues, alongside the corresponding BLT storage microplate logistics, i.e. heat sealing (see **FIG. 8A** and **8B**). Traditional tip-based liquid handlers allow source sample resuspension prior to its transfer (via pipetting), but come at the cost of reduced liquid handling accuracy at the nanoliter scale. Tip-less (acoustic) liquid handlers are tailored to perform highly accurate nanoliter scale liquid transfers, but do necessitate alternative means of source sample resuspension due to the tip-less mode of source sample transfer (via an acoustic wave pulse ejecting source samples from specialized

source microplates into inverted destination microplates).

We initially sought to address the BLT sedimentation by introducing high concentration of glycerol into the BLT storage buffer (see **MATERIALS**). This modification however proved insufficient in fully mitigating the BLT sedimentation issue. Ultimately, we found that additional quick BLT storage microplate orbital shaking (1500 rpm, 30 sec, 1.2 mm orbit; using an integrated, magnetic beads-compatible BioShake 5000-T elm from QInstruments), preceding BLT nanoliter dispenses into each destination microplate, allowed us to fully mitigate the issue (see **FIG. 8C**).

CONCLUSIONS

Cost- and time-efficient NGS library preparation and downstream DNA sequence verification of microbial genomes and newly assembled plasmid DNA are vital components of high-throughput genetic engineering and screening pipelines for synthetic biology and beyond.

The fully automated NGS library prep workflow outlined herein demonstrates how Ginkgo's RAC platform, equipped with state-of-the-art lab automation equipment, can be leveraged to substantially (10-fold) reduce per sample reagent costs and minimize scientists' hands-on time. We also show that additional adaptation of previously manual or partially manual biological workflows is typically necessary to fully realize significant laboratory automation gains.

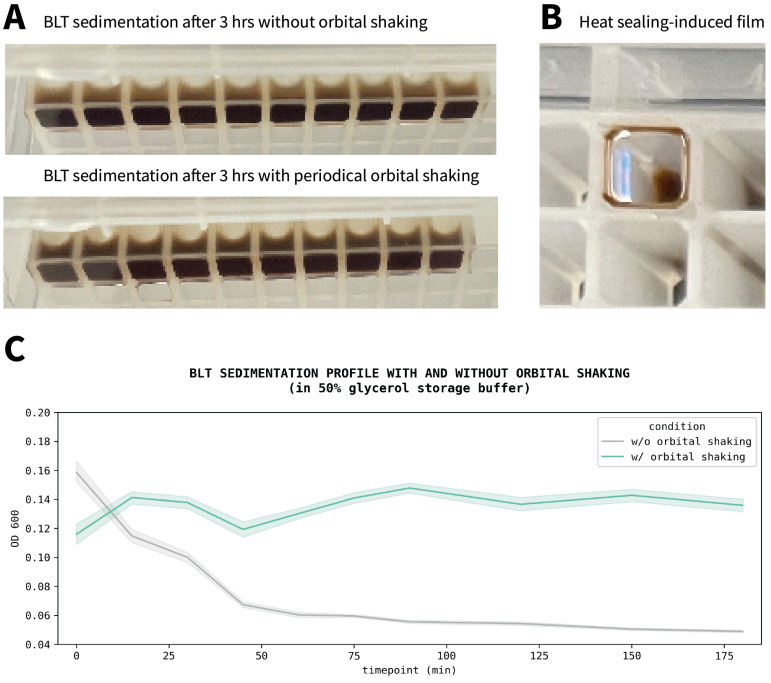


FIGURE 8. BLT Handling Challenges in a High-Throughput Environment; A BLT sedimentation in Echo Qualified 384-well Polypropylene Microplates after ~3 hrs without orbital shaking and with periodical orbital shaking (before each destination microplate liquid transfer); **B** Problematic BLT aliquoting residue- and heat-sealing-induced film formation above BLT-containing microplate wells, leading to "silent" BLT acoustic dispense failures - we avoided heat sealing in our workflow thereafter (no BLT evaporation was observed); **C** BLT sedimentation profiles with and without orbital shaking (BLT concentrations measured via a plate reader OD 600 measurement, and N=60 samples 95% confidence intervals highlighted)

REFERENCES

1. Pillay, S., San, J.E., Tshiabuila, D., Naidoo, Y., Pillay, Y., Maharaj, A., Anyaneji, Ugochukwu J, Wilkinson, E., Tegally, H., Lessells, R.J., Baxter, C., Oliveira, de and Giandhari, J. (2023). Evaluation of miniaturized Illumina DNA preparation protocols for SARS CoV-2 whole genome sequencing. *PLOS ONE*, [online] 18(4), pp.e0283219. Available at: https://doi.org/10.1371/journal.pone.0283219.

2. Shapland, E.B., Holmes, V., Reeves, C.D., Sorokin, E., Durot, M., Platt, D., Allen, C., Dean, J., Serber, Z., Newman, J. and Chandran, S. (2015). Low Cost, High-Throughput Sequencing of DNA Assemblies Using a Highly Multiplexed Nextera Process. *ACS Synth. Biol.*, [online] 4(7), pp.860–866. doi:https://doi.org/10.1021/sb500362n.

3. Suckling, L., McFarlane, C., Sawyer, C., Chambers, S.P., Kitney, R.I., McClymont, D.W. and Freemont, P.S. (2019). Miniaturisation of high-throughput plasmid DNA library preparation for next-generation sequencing using multifactorial optimisation. *Synthetic and Systems Biotechnology*, [online] 4(1), pp.57–66. doi:https://doi.org/10.1016/j.synbio.2019.01.002.

MATERIALS

- Four (4) 96-well microplates with each well pre-filled with technical replicate saturated cell culture *S. cerevisiae* CEN.PK strain 5 OD 600 cell pellets
- Three (3) 96-well microplates with each well pre-filled with technical replicate purified 1 ng / µL pUC19 plasmid DNA
- One (1) 96-well microplate with each well pre-filled with nuclease-free water (negative control)
- Yeast cell lysis mix A (100 mM potassium L-glutamate, 10 mM magnesium L-glutamate, 1% w/v polyvinylpyrrolidone, 50 mM HEPES, pH 7.5, 0.05 mg / mL RNase A from Lucigen, 0.05 U / µL zymolyase from Zymo Research)
- Yeast cell lysis mix B (10 µg / µL proteinase K from Lucigen)
- Bead-linked Transposomes (BLTs) from the Illumina DNA Prep kit, buffer-exchanged with a buffer mitigating BLT sedimentation (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.09% v/v Triton X-100, 50% v/v glycerol, 0.5 mM DTT, 0.2 mM EDTA, pH 8.0)
- Tagment Buffer 1 (TB1) from the Illumina DNA Prep kit
- 0.5% v/v SDS
- PCR master mix (1X KAPA HiFi HotStart ReadyMix from Roche, 1 M betaine, 0.5 µM fwd CDI indexing primer, 0.5 µM rev CDI indexing primer)