

For: Avian and Reptilian Developmental Biology, Methods in Molecular Biology [Guojun Shen, Editor]

Comparative genomics as a foundation for evo-devo studies in birds

Phil Grayson, Simon YW Sin, Timothy B Sackton, and Scott V Edwards

Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology,

Harvard University, Cambridge, MA 02138 USA

Corresponding author:
Scott V. Edwards
Museum of Comparative Zoology
Harvard University
26 Oxford Street
Cambridge, 02138
617-384-8082
sedwards@fas.harvard.edu

Running head: Avian comparative genomics

Summary

Developmental genomics is a rapidly growing field and high quality genomes are a useful foundation for comparative developmental studies. A genome streamlines and simplifies the development of primers used to isolate putative regulatory regions for enhancer screens, cDNA probes for *in situ* hybridization, micro RNAs (miRNAs) or short hairpin RNAs (shRNA) for RNA interference (RNAi) knockdowns, mRNAs for misexpression studies, and in recent years, even guide RNAs (gRNAs) for CRISPR knockouts. A high quality genome also forms an essential reference onto which the data from numerous assays and experiments, including ChIP-seq, ATAC-seq, and RNA-seq, can be mapped. Finally, much can be gleaned from comparative genomics alone, including identification of highly conserved putative regulatory regions. This chapter provides an overview of laboratory and bioinformatics protocols for DNA extraction, library preparation, library quantification and genome assembly, from fresh or frozen tissue to a draft avian genome. Generating a high quality draft genome can provide a developmental research group with excellent resources for their study organism, opening the doors to many additional assays and experiments.

Key Words: Genome assembly, library preparation, avian, bird, developmental genomics, ALLPATHS-LG, comparative genomics.

1. Introduction

1.1 Avian Developmental and Genomic Resources

A long history of study has made the chicken into an important model for analyzing developmental patterns and processes across diverse time scales and cell types [1]. Avian developmental biology is not limited to chicken: developmental staging systems for quail, zebra finch, and emu are also available [2–4]. Arguably one of the most important advancements for developmental or molecular avian research was the sequencing of the chicken (*Gallus gallus*) genome in 2004 [5]. In the ten years that followed, more than 50 additional avian genomes have been published, most of them with the December 2014 release of the *Science* Special Issue, “A Flock of Genomes”, examining data from the genomes of 48 birds [see 6, 7]. The last two years have seen an increasing rate of avian genome sequencing; at the time of writing the total number of draft bird genomes available in public databases has soared to over 70. With an eventual goal of generating draft genomes of all 10,500 species of living birds, the Bird 10,000 Genomes project (B10K) has promised to increase this total even more rapidly [8].

1.2 Avian Genomes

Avian genomes are relatively small, around 1-1.26 Gb, and low in repetitive elements compared to other amniotes [6, 9]. These two factors make achieving high-quality reference genomes relatively straightforward for research groups with the necessary protocols and infrastructure, including a few key instruments for library preparation and quantification, and access to Illumina sequencing machines. In the last several years our laboratory has produced a number of high quality draft genomes of birds. Each genome has required approximately 2 full lanes of Illumina HiSeq 2500 v4 high output sequencing to achieve reasonable coverage and quality. The ALLPATHS-LG assembly approach [10] used by our group made a strong showing during the

Assemblathon 2 comparative study of assembly methods and has since been applied frequently across diverse vertebrate taxa [11]. Sequencing technology is rapidly changing, as discussed in Section 1.4 below, and while the strategy outlined here has proved to be robust and cost-effective in our hands, new approaches will likely surpass the ALLPATHS-LG/Illumina strategy in both cost and quality in the near future.

1.3 ALLPATHS-LG and Basic Sequencing Strategies

The ALLPATHS-LG assembly algorithm requires at minimum a single overlapping fragment library (e.g., 220 bp library sequenced using 2x125 bp paired end reads), as well as a jumping library (e.g., 3 kb Nextera mate pair library, also sequenced using 2x125 bp reads). The jumping library allows the assembler to “jump” over stretches of repetitive DNA because each end of a mate pair read provides linkage information about bases that appear approximately 3 kb from one another in the genome. In the preparation of the jumping library, DNA is “tagmented” (fragmented and tagged with junction adapters in one step using transposase) and then circularized (junction adapter to junction adapter), such that DNA on either side of the junction (the mate pairs) originated approximately 3 kb away on the same strand in the native genomic DNA. The circularized DNA is then fragmented a second time and only junction adapter-containing fragments are collected and used to produce the library. For genomes that are larger and more complex than those of birds, a larger jumping library (e.g., 6 kb) or a fosmid jumping library (e.g., 40kb) can be incorporated to span long stretches of repeats [10].

Here, we report basic assembly statistics from 17 unpublished draft avian genomes generated in our laboratory to serve as a guide when planning the sequencing depth per library and total coverage for a new avian ALLPATHS-LG project (Table 1). Genomes from our laboratory were sequenced exclusively on Illumina HiSeq 2500s using 2x125bp reads from Illumina HiSeq SBS

V4 high output sequencing runs alone, or in combination with a small proportion of reads (generated during test runs) using rapid run mode at 2x150bp. Multiplexing and sequencing was carried out with the aim of providing between 500 million and 1 billion paired-end reads (between 1-2 lanes of a v4 flowcell) per genome. Because sequencing comprises the largest portion of the total cost in generating a draft genome, the genome of the Wandering Albatross (*Diomedea exulans*), at only 356 million reads and 28.3X coverage, cost approximately one half of the higher coverage genomes (e.g., the emu (*Dromaius novaehollandiae*)). [Table 1 near here]

Scaffold and contig N50 scores are common metrics of genome quality, defined as the length of the element (scaffold or contig) above which 50% of the genome exists in elements of that length or longer. Following assembly, each contig contains mostly-contiguous DNA sequence. Contigs are linked together with stretches of unknown bases in between to generate scaffolds. Reads from jumping libraries span regions that are difficult to sequence or assemble, aiding in the assembly of both contigs and scaffolds. Some trends in genome quality that emerge from analyzing the 17 avian genome assemblies include:

1 – Contig N50 correlates with total number of raw reads (both jumping and fragment) (Fig. 1).

[Fig 1 near here]

2 – Scaffold N50 also correlates with total raw read number, but the correlation is less strong than for contig N50 (Fig. 2). [Fig 2 near here] Two groups emerge in this graph, one that appears to benefit from deeper sequencing, and one that does not (assemblies from the latter group have scaffold N50 values below 2 Mb). Despite high sequencing depth in the two kiwi and two albatrosses near the bottom right of the figure, the scaffold N50 is low. Quality of the input DNA, properties of the genomes (e.g., repeat content or types), or other factors could explain this pattern.

3 – We note several caveats regarding increasing sequencing depth as a means to achieve increased quality of assembly. Our laboratory has experimented with increasing sequencing depth without creating additional libraries in two instances, once for a very low coverage assembly of the Waved Albatross (*Phoebastria irrorata*) and once as an attempt to improve the 67X coverage assembly of the Black-footed albatross (*Phoebastria nigripes*) (Table 1). For the Waved albatross, the 1st assembly was produced with approximately 21X coverage, with contig and scaffold N50 values of 3.06×10^4 and 7.87×10^5 bp, respectively. By increasing the coverage to over 40X, this assembly was greatly improved (see Table 1). However, for the Black-footed albatross, nearly doubling the coverage to approximately 110X apparently did little to improve the assembly (contig and scaffold N50 values of 8.64×10^4 and 1.51×10^6 bp, respectively). This result suggests that there may be diminishing returns to higher coverage sequencing, at least in some cases. For example, if the input DNA is not of very high quality the diversity of the resulting libraries may be low. Mapping raw sequencing reads back to the assembled genome could provide information pertaining to library diversity to determine if additional sequencing of the current libraries could improve an assembly [12].

4 – Both our experience and the theory behind genome assembly suggest that the best way to improve scaffold size and N50 in an assembly is to generate an additional jumping library. Only one genome in our laboratory, that of the Leach's storm petrel (*Oceanodroma leucorhoa*), has been sequenced with both a 3 kb and 6 kb jumping library, and it has a scaffold N50 approximately twice as long as any of our other genomes, suggesting that an additional jumping library can greatly improve assembly quality (Table 1). The petrel assembly also has a slightly higher contig N50 compared to the other 16 genomes.

1.4 Alternative approaches to sequencing avian genomes

At the time of writing, the Illumina/ALLPATHS-LG strategy we describe here is likely the most widely available and cost-effective method for producing high quality draft avian genomes (Fig. 3) [Fig 3 near here]. However, new and emerging sequencing technologies have the potential to produce better, more contiguous assemblies for a similar, or cheaper, cost. Oxford Nanopore and Pacific Biosciences both have the advantages of single-molecule long reads, but still suffer from higher costs and error rates compared to the shorter Illumina reads [13, 14]. There is little doubt that deep sequencing on either of these platforms would provide the best possible assembly, but might cost up to 10 times more than the Illumina/ALLPATHS-LG approach described here [15]. Costs and error rate continue to rapidly decline on these platforms, and they may achieve parity with Illumina in the near future.

Another recent alternative is the Chromium library preparation system from 10X Genomics, which uses microfluidics to partition genomic DNA and label individual strands with over one million unique barcodes prior to sequencing on the Illumina platform [16]. By referring back to the original barcodes on the unfragmented DNA, the assembly algorithm can identify the original native DNA molecule from which a particular read is derived, thus generating haplotype-phased genomes with high scaffold and contig N50. The additional costs of the library prep for this approach compared to that for standard Illumina are relatively modest, but the efficiency and performance of the algorithms and methods in non-human contexts is not yet clear [17].

1.5 Draft genomes as foundations for developmental genomics

The first steps in sequencing a genome are the most critical. The high cost of errors during library preparation or quantification often places genome projects out of reach for many research laboratories. Our goal for this chapter was to provide an overview of an accessible method for

generating draft avian genomes, allowing laboratories to take advantage of new approaches to developmental biology allowed by comparative genomics (Fig. 4). [Fig 4 near here]

The identification of regulatory regions is of key importance for many developmental and evolutionary investigations. Comparative genomic studies have identified conserved non-exonic elements (CNEEs), stretches of noncoding sequence that are conserved across divergent taxa and evolving more slowly than a benchmark neutral rate, usually comprised of four-fold degenerate sites of protein coding regions. These CNEEs, which are often identified through the phastCons program following whole genome alignment [18], appear to serve developmental regulatory roles. When tested in a transgenic mouse enhancer assay, 115 out of 231 of the highly conserved human genomic regions drove reporter gene expression in various murine tissues, suggesting conserved enhancer activity [19].

Because of the important role of CNEEs in development, many computational methods have been used to predict their origins, functions, and the effect of repeated losses of orthologous elements across diverse taxa [18, 20–23]. CNEE discovery techniques have proven to be most powerful when coupled with genomic assays for tissue-specific regulatory activity. Using a 60-way vertebrate alignment for CNEE identification, alongside chromatin immunoprecipitation (ChIP)-seq data for active enhancer marks (histone 3 lysine 27 acetylation or H3K27ac) from developing mouse limbs, Booker et al. were able to identify and experimentally validate novel regulatory regions (referred to as Bat Accelerated Regions or BARs due to their rapid evolutionary rate in bat lineages) implicated in the development of the bat wing and conduct comparative *in situ* hybridization experiments on genes neighboring BARs of interest [24]. Since CNEEs are often found clustered near key developmental genes [20], interesting patterns of CNEE evolution can also aid in the generation of candidate gene lists for traits of interest.

Another recent paper on development and evolution of bat limbs used a similar comparative genomics approach to identify additional BARs [25]. These authors then examined comparative RNA-seq, *in situ* hybridization and both H3K27me3 (repressive mark) and H3K27ac (active mark) ChIP-seq data across three developmental stages to identify regulatory changes and expression differences for RNAs (both coding and long noncoding) implicated in shaping bat forelimb development through differential mesenchymal condensation.

In pigeons, comparative genomics, cross-strain breeding and developmental experiments alongside ChIP-seq from developing limbs allowed researchers to suggest that genes implicated in the origins of feathered feet in multiple pigeon breeds drive an alteration of limb identity, with the developing feathered hindlimbs displaying a localized shift toward forelimb-specific expression patterns of *Tbx5* and *Pitx1* through cis-regulatory changes [26]. Furthermore, the authors identified similar *Tbx5* expression patterns in the developing hindlimbs of convergently-feathered chicken breeds [26]. In other *Tbx5* research, an assay that identifies stretches of open chromatin from live intact cells (ATAC-seq) uncovered a CNE present in humans, bony fish and cartilaginous fish that appears to serve an essential role in the evolution of paired appendages across jawed vertebrates [27]. Injection of this enhancer region alongside *tbx5a* into finless zebrafish (*Danio rerio*) embryos rescued the finned phenotype, while injecting *tbx5a* alone was ineffective [27]. All of the developmental studies described above used comparative genomics alongside cutting-edge genomic and epigenomic assays for their model and trait of interest, illustrating the utility of one or more draft genomes for driving developmental research.

1.6 Overview

The methods section begins with avian tissue and ends with draft de novo genome assemblies (Fig. 5). [Fig 5 near here] Each section follows one or more published protocol(s) from the kits

and instruments recommended; DNA extraction (3.1) follows the protocol from the E.Z.N.A. Tissue DNA Kit. Fragment Library Preparation (3.2) follows protocols from the Qubit dsDNA HS Assay Kit, PrepX ILM 32i DNA Library Kit, IntegenX PCR Cleanup 8 Protocol, and the Covaris S220 instrument. Jumping (Mate Pair) Library Preparation (3.3) follows protocols from the Qubit, Covaris, Nextera Mate Pair Sample Preparation Kit, and Pippin Prep instruments. Library Quantification and Quality Control (3.4) follows protocols from Qubit, the High Sensitivity DNA Kit for the Agilent Bioanalyzer, and the Kapa Library Quantification Kit for Next Generation Sequencing. ALLPATHS-LG Assembly (3.5) utilizes commands from Fastq, cutadapt, TrimGalore!, and ALLPATHS-LG manuals. Notes (section 4) interspersed throughout the materials and methods provide tips and best practices that have benefitted our laboratory throughout these protocols.

This chapter serves as a general guide to achieving high quality draft genomes of birds, but we recommend referring to the protocols included in the kits themselves or updated online because companies often modify their protocols as they change reagents or optimize key steps.

2. Materials

2.1 DNA Extraction and Quality Control

1. Tissue sample (*see Note 1*).
2. DNA extraction kit such as E.Z.N.A. Tissue DNA Kit (cat. No. D3396, Omega bio-tek).
3. Standard 1.7 mL microcentrifuge tubes.
4. RNase A.
5. 100% ethanol.
6. 100% isopropanol.
7. Sterile scalpel blades.

8. Agarose.
9. 10X TAE buffer (cat. No. 15558042, ThermoFisher).
10. 1kb DNA Ladder and 6X loading dye (cat. No. N3232, NEB).
11. Agarose gel electrophoresis system.
12. Gel imager.
13. SYBR Safe DNA Gel Stain (cat. No. S33102, ThermoFisher).

2.2 Fragment Library Preparation

1. DNA from 2.1.
2. Qubit Fluorometer.
3. Qubit dsDNA HS Assay Kit (cat. No. Q32854, ThermoFisher Scientific).
4. Qubit assay tubes (cat. No. Q32856, Life Technologies) or Axygen PCR-05-C tubes (cat. No. 10011-830, VWR).
5. Apollo 324 NGS Library Preparation System (*see Note 2*).
6. Apollo consumables (*see Note 3*): PCR strip-tubes, microtiter plates, filter tips, piercing tips, PCR strip-caps, reservoirs, and 1.1 mL tubes (cat. No. 300019, 300026-300029, 300031 and 300033 respectively).
7. PrepX ILM 32i DNA Library Kit (*see Note 2*) (cat. No. 400076, Wafergen).
8. PrepX Complete ILMN DNA Barcodes (cat. No. 400077-1 to 400077-5, Wafergen).
9. Covaris S220 Shearing Device and microTUBE holder.
10. Covaris microTUBE (6 x 16 mm) glass tubes (cat. No. 520045, Covaris).
11. 70% EtOH.
12. 2.5 M NaCl.

13. UltraPure DNase/RNase-free distilled water (cat. No. 10977023, ThermoFisher Scientific). This will be referred to as water throughout.

2.3 *Jumping [Mate Pair] Library Preparation*

1. Nextera Mate Pair Sample Preparation Kit (cat. No. FC-132-1001, Illumina).
2. Qubit Fluorometer.
3. Qubit dsDNA HS Assay Kit (cat. No. Q32854, ThermoFisher Scientific).
4. Qubit assay tubes (cat. No. Q32856, Life Technologies) or Axygen PCR-05-C tubes (cat. No. 10011-830, VWR).
5. Covaris S220 Shearing Device and microTUBE holder.
6. Covaris T6 (6 x 32 mm) glass tubes (cat. No. 520031, Covaris).
7. Covaris Snap-Cap (8 mm) Teflon Silicon Septa (cat. No. 520042, Covaris).
8. Pippin Prep.
9. Pippin Prep 0.75% Agarose Gel Cassettes with Ethidium Bromide (cat. No. CSD7510, Sage Science)
10. Maxymum Recovery 1.7 mL microcentrifuge tubes (cat. No. MCT-175-L-C, Axygen).
11. Genomic DNA Clean and Concentrator Kit (cat. No. D4010 or D4011, Zymo Research).
12. AMPure XP Beads (cat. No. A63880, Beckman Coulter).
13. Dynabeads M-280 streptavidin magnetic beads (cat. No. 112-05D, Invitrogen).
14. Magnetic rack for 1.7 mL microcentrifuge tubes; MagnaRack for example (cat. No. CS15000, Invitrogen).

2.4 *Library Quantification and Quality Control*

1. Agilent 2100 Bioanalyzer.
2. High Sensitivity DNA Kit for Bioanalyzer (cat. No. 5067-4626, Agilent Technologies).

3. Kapa Library Quantification Kit for Next Generation Sequencing (cat. No. KK4824, Kapa Biosystems). Catalog number above is for the universal kit.
4. qPCR thermocycler.
5. Preferred diluent (10 mM Tris-HCl, pH 8.0 and optional 0.05% Tween 20).
6. qPCR plates or tubes with clear sealing film or lids.

2.5 ALLPATHS-LG Assembly

1. Computer with Fastqc installed. Available here:
<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
2. Computer with cutadapt installed. Available here:
<http://cutadapt.readthedocs.io/en/stable/installation.html>
3. Computer with Trim Galore! installed. Available here:
http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
4. High memory (>256 Gb) server running ALLPATHS-LG (*see Note 4*). Available here:
<http://software.broadinstitute.org/allpaths-lg/blog/>

3. Methods

3.1 DNA Extraction

1. This section follows the protocol from the E.Z.N.A. Tissue DNA Kit (*see Note 5*).
2. Preheat blocks or water baths to 55°C and 70°C, place elution buffer at 70°C, and add appropriate volume of 100% ethanol and 100% isopropanol to DNA wash buffer and HBC buffer respectively (varies with kit size) before beginning.
3. Weigh and mince 30 mg of tissue. Place in 1.7 mL microcentrifuge tube.
4. Add 220 µL TL Buffer and flick-mix by hand. Do not vortex at any point (*see Note 6*).
5. Add 25 µL OB Protease Solution and mix.

6. Incubate at 55°C with agitation, either constant shaking or frequent flick-mixing every 20-30 minutes (*see Note 7*).
7. Once tissue is digested, add 4 µL RNase A (100 mg/mL) per 30 mg tissue and incubate for 2 minutes at room temperature. This will remove RNA from your sample and allow you to better determine the quality of your DNA sample on an agarose gel.
8. Centrifuge at maximum speed (use max speed for all DNA extraction spins) for 5 minutes and transfer supernatant to sterile 1.7 mL microcentrifuge tube.
9. Add 220 µL BL Buffer and mix thoroughly.
10. Incubate at 70°C for 10 minutes.
11. Add 220 µL of 100% ethanol and mix thoroughly.
12. Transfer full volume to HiBind DNA Mini Column placed in a collection tube.
13. Centrifuge at max speed for 1 minute.
14. Discard filtrate and replace collection tube.
15. Add 500 µL HBC buffer (ensure that 100% isopropanol has been added).
16. Centrifuge at max speed for 30 seconds.
17. Discard collection tube and filtrate. Apply new collection tube to column.
18. Add 700 µL DNA wash buffer (ensure that 100% ethanol has been added).
19. Centrifuge at max speed for 30 seconds.
20. Discard filtrate and replace collection tube.
21. Repeat 700 µL DNA wash buffer steps one additional time.
22. Centrifuge empty column and collection tube for 2 minutes to remove excess wash buffer.

23. Transfer column to new collection tube and centrifuge for 1 additional minute to remove excess wash buffer.
24. Transfer HiBind DNA Mini Column into sterile nuclease-free 1.5 mL microcentrifuge tube.
25. Add 100-200 heated 70°C elution buffer to column. Incubate columns at 70°C for 2 minutes.
26. Centrifuge for 1 minute. Second elution can be carried out with additional elution buffer.
27. Make 1X TAE stock from 10X TAE buffer using dH₂O.
28. Based on total volume of 1X TAE used, weigh and mix in appropriate amount of agarose to create 0.5%-1% agarose gel. Microwave and mix solution until homogeneous and cool until comfortably warm to the touch. Add SYBR Safe (approximately 1ul per 10 mL of gel solution). Pour, harden, load, and run gel at 120 V.
29. When the gel is imaged, samples should show single band with little to no streaking underneath (Fig. 6). [Fig 6 near here]

3.2 Fragment Library Preparation

1. This section follows protocols from: Qubit dsDNA HS Assay Kit, PrepX ILM 32i DNA Library Kit, IntegenX PCR Cleanup 8 Protocol, and the Covaris S220 instrument
2. Following the Qubit dsDNA HS Assay Kit protocol, quantify DNA prior to library preparation. The full protocol is included below.
3. Label tops of assay tubes for samples and 2 standards.
4. Calculate and mix the appropriate volume for Qubit working solution for all samples and standards in a plastic microcentrifuge or Falcon tube. The working solution is composed of Qubit dsDNA HS Reagent diluted 1:200 in Qubit dsDNA HS Buffer. Each sample

requires 199 μL of working solution (if the readings are too low using 199 μL , the Qubit protocol allows volumes as low as 180 μL for samples) and the two standards both require 190 μL .

5. Each tube will be brought to 200 μL with standard or sample, vortexed and incubated at room temperature for two minutes before reading (*see Note 8*).
6. On the Qubit, select DNA, then dsDNA High Sensitivity.
7. Read standard 1, then standard 2.
8. Read sample and calculate concentration on the Qubit screen based on volume of sample used (e.g., 1 μL sample with 199 μL of working solution).
9. Following the Waforgen Biosystems PrepX ILM 32i DNA Library Kit protocol, the DNA concentration working range is 0.067–6.67 ng/ μL . All our libraries were generated near the top end of this range.
10. Turn on Covaris and water bath at least 30 minutes prior to use.
11. Fill water tank with distilled water between the 10 and 15 on the fill scale, which should result in a reading between 10 and 15 on the run scale once the apparatus is lowered into the tank and water is displaced.
12. Turn the machine on and open the Sonolab software.
13. Allow the water to degas and reach optimal temperature (e.g., 6°C).
14. To fragment DNA, pipette 50–130 μL of sample through the pre-slit cap on a Covaris microTUBE.
15. Place microTUBE in microTUBE holder and place holder in Covaris (*see Note X*).
16. For 220 bp insert size, set the following protocol on the Sonolab software: Duty Cycle/Duty Factor: 10%, Intensity: 4 (Peak Power Intensity on the S220 is set to 140),

Cycles/bursts: 200, Time/cycle: 9 seconds, Number of cycles: 8, Total process time: 72 seconds.

17. Press run. Following fragmentation, remove sample and place on ice.
18. Prior to Apollo setup, anneal universal Wafergen adapters and barcodes in their plate using a 2-step thermocycler protocol (95°C for 5 minutes, 70°C for 15 minutes).
19. Remove plate from thermocycler and cool on bench. Add 6 µL of water to each well and mix. Reseal plate after use and store at -20°C. This is only done once (not each time a library preparation is carried out).
20. To utilize adapters, mix 2 µL of adapter (from the plate above) with 13 µL of water for each Apollo run (this is the Adapter Mix). Keep on ice.
21. To make Apollo Ligation Mix (per 4 volumes), combine 52.8 µL of PrepX Ligase Buffer, 4.4 µL of PrepX Ligase Enzyme, 8.8 µL of water. Keep on ice.
22. Turn on Apollo and select Library Prep, DNA, ILM, 220 bp.
23. Follow the on-screen instructions for setup of Apollo consumables.
24. Once the “Cooling” notification disappears, load the following in PCR strip-tubes as instructed on-screen: 15 µL of Covaris-sheared sample, 15 µL of Adapter Mix, 64 µL of Ligation Mix (per 4 reactions).
25. Load 225 µL of 2.5 M NaCl (supplied by Wafergen) to the Apollo microtiter plate as instructed on-screen (per 4 reactions).
26. Add water and 70% ethanol to reservoirs 1 and 3 as instructed on-screen.
27. Load the enzyme strip(s) and place retention plates on blocks 3 and 4 as instructed on-screen.
28. Close the instrument door and press “Run”. The run will end in 3 hours, 20 minutes.

29. Retrieve products from block 3, rows 9-12. The libraries should be ~ 15 μL .
30. Qubit libraries as above and prepare the PrepX PCR reaction mixture (25 μL of PrepX PCR Master Mix, 2 μL of PrepX PCR Primers, 5 ng of Apollo library and water to make up the total volume to 50 μL).
31. Run the following PCR protocol: 98°C for 30 seconds followed by 5 cycles of 98°C for 10 seconds, 60°C for 30 seconds, and 72°C for 30 seconds. Following the 5 cycles, hold at 72°C for 300 seconds and hold at 4°C until removed from the thermocycler.
32. Proceed to PCR Cleanup 8 Protocol on the Apollo. Warm AMPure Beads to room temperature and vortex thoroughly just prior to running protocol.
33. Turn on the Apollo. Press “Utility Apps” and “PCR Cleanup 8”.
34. Place empty consumables, 6 mL of 100% ethanol, and 6 mL of water on the Apollo deck as instructed on screen.
35. Once “Cooling” has disappeared, place strips containing 50 μL of sample and those containing 50 μL of freshly-vortexed AMPure beads on Apollo deck as instructed on screen.
36. Press “Run”.
37. After 16 minutes, collect clean products from Block 3, Row 4 and place on ice.
38. Bring 10 μL of clean library up to 15 μL with water.
39. Proceed to library quantification and quality control.

3.3 *Jumping [Mate Pair] Library Preparation*

1. This section follows protocols from the Nextera Mate Pair Sample Preparation Kit and the Qubit, Covaris, and Pippin Prep instruments.

2. For the Gel-Plus protocol (using the Pippin Prep), the Nextera Mate Pair Library Prep Reference Guide recommends 4 µg of input DNA.
3. Quantify DNA with Qubit as described in section 3.2.
4. Thaw the Tagment Buffer Mate Pair on ice.
5. Set a heat block to 55°C.
6. Add 4 µg of gDNA to 1.7 mL microcentrifuge tube (max volume 308 µL).
7. Make up remainder of 308 µL with water.
8. Add 80 µL of Tagment Buffer Mate Pair.
9. Add 12 µL of Mate Pair Tagment Enzyme.
10. Flick-mix (do not vortex) and collect volume at bottom of tube using minicentrifuge.
11. Incubate for 30 minutes at 55°C.
12. Purify the tagmentation reaction using the Zymo Genomic DNA Clean & Concentrator kit. Begin by adding 2 volumes (2x400=800 µL) of Zymo ChIP DNA Binding Buffer to the tagmentation reaction. Pipette to mix.
13. Transfer up to 800 µL of mixture to Zymo-Spin IC-XL column in a collection tube.
14. Centrifuge at top speed for 30 seconds. Discard flow-through.
15. Add remaining reaction to column and repeat previous two steps.
16. If this is the first time using the Zymo kit, add 24 mL 100% ethanol to the 6 mL (or 96 mL 100% ethanol to the 24 mL) DNA Wash Buffer concentrate.
17. Wash two times by adding 200 µL of Zymo DNA Wash Buffer to the column, spinning at top speed for 1 minute and discarding the flow-through.

18. Centrifuge the empty column at top speed with the lid open for 1 minute to remove excess wash buffer. Discard the collection tube and transfer the column to a new Axygen Maxymum Recovery 1.7 mL microcentrifuge tube (*see Note 10*).
19. Add 30 μ L of Resuspension Buffer (RSB from Nextera Kit) and incubate at room temperature for 1 minute before centrifuging for 1 minute at full speed to elute purified DNA (*see Note 11*).
20. DNA can be stored at -20°C for up to 24 hours at this step, but to complete the protocol in two days, proceed to the next step (*see Note 12*).
21. Thaw 10x Strand Displacement Buffer and dNTPs at room temperature and place on ice when completely thawed.
22. Set heat block to 20°C.
23. To the 30 μ L of purified tagmented DNA sample, add 132 μ L water, 20 μ L of 10x Strand Displacement Buffer, 8 μ L of dNTPs, and 10 μ L of Strand Displacement Polymerase.
24. Flick-mix and collect volume at bottom of tube using minicentrifuge.
25. Incubate for 30 minutes at 20°C. During this time, remove AMPure XP beads from 4°C and allow them to reach room temperature.
26. Purify Strand Displacement mix using AMPure XP beads. Begin by vortexing beads thoroughly (vortex beads often to ensure they are evenly dispersed).
27. Prepare 800-900 μ L of 70% ethanol per library preparation.
28. To the 200 μ L of strand displaced DNA, add 100 μ L of freshly-vortexed AMPure XP beads.
29. Flick-mix and collect volume at bottom of tube using minicentrifuge.

30. Incubate at room temperature for 15 minutes repeating the previous step every 2 minutes.
31. Place on magnetic rack for 5-10 minutes.
32. Leave tube on rack to remove and discard all supernatant without disturbing beads.
33. Leave tube on rack for 2 washes with 400 μ L of 70% ethanol, removing and discarding supernatant after 30 seconds (*see Note 13*).
34. Once all supernatant has been removed, allow the beads to dry on the magnetic rack for 10-15 minutes. Small volumes of supernatant can be collected manually with a 10 μ L pipette tip during the drying step.
35. Add 30 μ L of RSB and remove tube from magnetic rack.
36. Flick-mix and collect volume at bottom of tube using minicentrifuge.
37. Incubate at room temperature for 5 minutes. Your sample is now in the supernatant.
38. Place on magnetic rack for 5 minutes.
39. Carefully collect supernatant and pipette to new 1.7 mL microcentrifuge tube.
40. DNA can be stored at -20°C for up to 7 days at this step, but to complete the protocol in two days, proceed to the next step.
41. Size selection can be carried out using the Sage Science 0.75% Cassette in the Pippin Prep for jumping libraries up to 8 kb. This will be described below. For jumping libraries 8-10 kb in size (which our lab has not utilized), size selection can be carried out with agarose gel electrophoresis, followed by a gel cleanup kit, or the BluePippin from Sage Science.
42. Turn on Pippin Prep. Click “Protocol Editor”.
43. Select “NEW”.

44. Select appropriate cassette (0.75%).
45. Select “End Run when Elution is Completed”.
46. Under “Reference Lane”, select a lane number for the marker D, then “APPLY REFERENCE TO ALL LANES”.
47. Select “Range” and supply values (e.g., 2-5 kb or 4-8 kb). If the Pippin Prep only allows tight cuts with the 0.75% cassette, supply your target size (e.g., 3 kb or 6 kb). Save the protocol.
48. Fill the rinse cassette with distilled water and place into the optical nest. Close the lid and wait 20 seconds.
49. Press CALIBRATE button on screen. This will prompt the message, “Calibration not done.”
50. Place the Calibration fixture dark side down over the 5 LED detectors on the rinse cassette.
51. Close the lid and ensure that 0.80 is in target setting.
52. Click CALIBRATE. If Calibration OK, press EXIT and proceed to samples.
53. The cassette is constructed as follows from left to right: Negative Buffer Chambers, Sample Wells, Elution Modules, Positive Buffer Chambers.
54. Inspect the 0.75% Cassette for damage to the gel and ensure that buffer reservoirs are a minimum of 50% full (add buffer if necessary). Do not use lanes that appear to be damaged.
55. Look at underside of cassette to ensure that there are no bubbles under the gel (between the gel and the plastic). Do not use lanes where gel bubbles appear to be present.

56. Tilt Positive Buffer Chambers up to gather bubbles near far right wall. Tap sides and bottom of cassette gently to dislodge any small bubbles between the Elution Modules and the Positive Buffer Chambers.
57. Place the cassette into the optical nest of the Pippin Prep while maintaining this position to ensure that bubbles do not get trapped between the Elution Modules and the Positive Buffer Chambers.
58. Place one hand on cassette and slowly remove adhesive strips by the white tabs provided.
59. Remove old buffer from elution modules and replace with 40 μ L of fresh electrophoresis buffer (*see Note 14*).
60. Seal the elution modules with adhesive tape strips provided.
61. Fill Sample Wells to maximum with buffer if any are not full.
62. Test continuity by closing the lid and pressing the TEST button on screen.
63. Once PASS appears on the screen, hit RETURN (*see Note 15*).
64. Combine entire 30 μ L of supernatant from AMPure bead purification with 10 μ L of room temperature loading solution.
65. Flick-mix and collect volume at bottom of tube using minicentrifuge.
66. Remove 40 μ L of buffer from Sample Well in target lane. Do not puncture agarose.
67. Add 40 μ L of sample to Sample Well.
68. Repeat for additional samples and for Marker D.
69. Begin run.

70. Once run is complete, carefully remove the adhesive strip over the Elution Modules one at a time and transfer size-selected product into new 1.7 mL microcentrifuge tube (*see Note 14*).
71. Purify the DNA from the Pippin Prep by once again using the Zymo Genomic DNA Clean and Concentration Kit, but note that the protocol different from above (*see Note 16*).
72. Pipette to mix 5 volumes of ChIP Binding Buffer into the Pippin Prep elution.
73. Transfer the mixture to a Zymo-Spin IC-XL column in a collection tube.
74. Centrifuge at top speed for 30 seconds. Discard flow-through.
75. Wash two times by adding 200 μ L of Zymo DNA Wash Buffer to the column, spinning at top speed for 1 minute and discarding the flow-through.
76. Discard the collection tube and transfer the column to a new 1.7 mL microcentrifuge tube.
77. Add 10 μ L of RSB directly to the column matrix and incubate at room temperature for 1 minute before centrifuging for 30 seconds at full speed to elute purified DNA.
78. DNA can be stored at -20°C for up to 24 hours at this step, but to complete the protocol in two days, proceed to the circularization step.
79. Run Qubit dsDNA HS Assay on samples as described in methods.
80. Utilize up to 600 ng of purified DNA for the circularization reaction (*see Note 17*).
81. Thaw Circularization Buffer at room temperature, then place on ice.
82. Set heat block to 30°C.

83. Using up to 600 ng of DNA from the previous step, add enough water to make the total volume 268 μL , then 30 μL of Circularization Buffer 10X, and 2 μL of Circularization Ligase.
84. Flick-mix and collect volume at bottom of tube using minicentrifuge.
85. Incubate overnight (12-16 hours) at 30°C. This is a good stopping point for day 1.
86. Set 2 heat blocks to 37°C and 70°C.
87. Thaw Stop Ligation Buffer and place on ice.
88. Add 9 μL of Exonuclease to overnight circularization reaction.
89. Flick-mix and collect volume at bottom of tube using minicentrifuge.
90. Incubate for 30 minutes at 37°C.
91. During this incubation, turn on and prepare the Covaris S220.
92. Fill the reservoir with distilled water and allow machine to degas.
93. Chill Covaris S220 water bath to 6°C.
94. Incubate samples from above for 30 minutes at 70°C.
95. During this incubation, set up Covaris S220 protocol.
96. Set Intensity to 8 (or Peak Power Intensity to 240 on the S220), Duty Cycle/Duty Factor to 20%, Cycles per Burst to 200, and Time to 40 seconds.
97. Flick-mix and add 12 μL of Stop Ligation Buffer to samples.
98. Flick-mix and collect volume at bottom of tube using minicentrifuge.
99. Transfer the entire volume of sample (~320 μL) to a Covaris T6 tube. Fill tube with water if space remains after addition of sample. Carefully place the Snap Cap onto tube and press firmly (*see Note 18*).
100. Snap the T6 tube into the tube holder and place it into the Covaris S220.

101. Shear the DNA and transfer to a new 1.7 mL microcentrifuge tube on ice.
102. Vortex and shake M-280 Streptavidin Magnetic Beads to resuspend.
103. Transfer 20 μ L of M-280 Streptavidin Magnetic Beads beads per library preparation to a new 1.7 mL microcentrifuge tube.
104. Place on magnetic rack for 1-2 minutes.
105. Remove and discard supernatant.
106. Wash the beads 2 times by adding 40 μ L of Bead Bind Buffer for every 20 μ L of original beads, incubating for 1-2 minutes and removing all supernatant.
107. Remove the beads from the magnetic rack and add 300 μ L of Bead Bind Buffer per 20 μ L of original beads. Mix thoroughly and spin down.
108. Add 300 μ L of beads in Bead Bind Buffer to 300 μ L of sheared DNA.
109. Incubate for 15 minutes at 20°C. Every 2 minutes during this incubation, flick-mix and collect volume at bottom of tube using minicentrifuge. Do not spin for more than a second to avoid pelleting the beads.
110. Following incubation, spin briefly in a minicentrifuge (5-10 seconds).
111. Place tube on the magnetic rack for 1-2 minutes. During this incubation, remove the End Repair Mix from the freezer and place on bench to thaw. Once thawed, move the End Repair Mix to ice.
112. Remove and discard supernatant from microcentrifuge tube.
113. Wash 4 times with Bead Wash Buffer by adding 200 μ L of Bead Wash Buffer to the beads, flick-mixing and spinning down for 1-2 seconds, placing the mixture on the magnetic rack for 30 seconds and removing all supernatant.

114. Wash 2 times with RSB by adding 200 μ L of RSB to the beads, flick-mixing and spinning down for 5-10 seconds, placing the mixture on the magnetic rack for 30 seconds and removing and discarding all supernatant. The supernatant from the second wash should not be removed until the End Repair Mix is thawed (from above) and the End Repair Mixture (in next step) has been made. Set a heat block to 30°C during this step.
115. For each library preparation, mix 40 μ L of End Repair Mix and 60 μ L of water to create the End Repair Mixture.
116. Remove all supernatant from tubes.
117. Spin briefly (approximately 5 seconds) on the minicentrifuge to collect additional supernatant.
118. Remove all excess supernatant with a 10 μ L pipette.
119. Add 100 μ L of End Repair Mixture to beads.
120. Flick-mix and collect volume at bottom of tube using minicentrifuge.
121. Incubate for 30 minutes at 30°C. During this incubation, place A-Tailing Mix on surface of ice in ice bucket (it will not thaw if placed into the ice).
122. Following incubation, spin in a minicentrifuge for 5-10 seconds.
123. Place the tube on magnetic rack for 1 minute. If the A-Tailing Mix has not thawed on ice, begin thawing at room temperature now.
124. Remove and discard supernatant.
125. Wash 4 times with Bead Wash Buffer by adding 200 μ L of Bead Wash Buffer to the beads, flick-mixing and spinning down for 5-10 seconds, placing the mixture on the magnetic rack for 30 seconds and removing all supernatant.

126. Wash 2 times with RSB by adding 200 μL of RSB to the beads, flick-mixing and spinning down for 5-10 seconds, placing the mixture on the magnetic rack for 30 seconds and removing and discarding all supernatant. The supernatant from the second wash should not be removed until the A-Tailing Mix is thawed (from above) and the A-tailing reaction mixture (in next step) has been made. Set a heat block to 37°C during this step.
127. For each reaction, mix 12.5 μL A-Tailing Mix and 17.5 μL of water to create the A-tailing reaction mixture.
128. As above, remove all supernatant from the tube, spin for 5-10 seconds and then remove excess using a 10 μL pipette.
129. Add 30 μL of A-tailing reaction mixture.
130. Flick-mix and collect volume at bottom of tube using minicentrifuge.
131. Incubate for 30 minutes at 37°C. During this incubation, thaw DNA Adapter Indexes and Stop Ligation Buffer at room temperature. Once thawed, place on ice. Select individual DNA Adapters for each library (*see Note 19*).
132. To the 30 μL A-tailing reaction mix, add 2.5 Ligation Mix, 4 μL water, and 1 μL of unique DNA Adapter Index. Record which adapter has been used for each library.
133. Flick-mix and collect volume at bottom of tube using minicentrifuge.
134. Incubate for 10 minutes at 30°C. During this incubation, place the Enhanced PCR Master Mix and PCR Primer Cocktail tubes at room temperature until thawed. Once thawed, place on ice.
135. Add 5 μL of Ligation Stop Buffer.
136. Spin for 5-10 seconds in a minicentrifuge.

137. Wash 4 times with Bead Wash Buffer by adding 200 μ L of Bead Wash Buffer to the beads, flick-mixing and spinning down for 5-10 seconds, placing the mixture on the magnetic rack for 30 seconds and removing all supernatant.
138. Wash 2 times with RSB by adding 200 μ L of RSB to the beads, flick-mixing and spinning down for 5-10 seconds, placing the mixture on the magnetic rack for 30 seconds and removing and discarding all supernatant. The supernatant from the second wash should not be removed until the Enhanced PCR Master Mix and PCR Primer Cocktail have thawed and the PCR reaction mixture (in next step) has been made.
139. For each library preparation, mix 20 μ L of Enhanced PCR Master Mix, 5 μ L of PCR Primer Cocktail, and 25 μ L of water to create the PCR reaction mixture.
140. Program a thermocycler with the following PCR program: 98°C for 30 seconds, followed by 10-15 cycles of 98°C for 10 seconds, 60°C for 30 seconds, and 72°C for 30 seconds (*see Note 20*). Following these cycles, hold at 72°C for 5 minutes and hold at 4°C.
141. As above, remove and discard supernatant, spin tube to collect excess supernatant, replace the tube on the magnetic rack, and remove excess with a 10 μ L pipette.
142. Add 50 μ L of PCR reaction mixture to the beads and pipette to mix.
143. Transfer the mixture to PCR tubes and run PCR program.
144. Following PCR, the library can be kept at -20°C for up to 7 days. This is a good stopping point for day 2.
145. Bring AMPure XP beads to room temperature.
146. Prepare 400 μ L of fresh 70%-80% ethanol per library.

147. Place PCR tubes on magnetic rack for 1 minute. If necessary, pipette library from PCR tube into new 1.7 mL microcentrifuge tube to use the same magnetic rack from previous steps.
148. Transfer 45 μ L of PCR product to a new 1.7 mL microcentrifuge tube
149. Add 30 μ L of freshly-vortexed AMPure XP beads to this volume.
150. Flick-mix and collect volume at bottom of tube using minicentrifuge.
151. Incubate at room temperature for 5 minutes.
152. Place on magnetic rack for 5 minutes.
153. Remove and discard supernatant.
154. Wash 2 times with 70%-80% ethanol by adding 200 μ L of ethanol to the beads, incubating for 30 seconds and removing the supernatant. If beads are disturbed, allow them to settle once more before removing supernatant.
155. Dry the beads for 10-15 minutes. Remove excess ethanol during this time with a 10 μ L pipette tip.
156. Remove the tube from the magnetic rack and resuspend beads in 20 μ L of RSB. Flick-mix and collect volume at bottom of tube using minicentrifuge.
157. Incubate at room temperature for 5 minutes.
158. Incubate on the magnetic rack for 5 minutes. The library is now in the supernatant.
159. Transfer the supernatant to a new 1.7 mL microcentrifuge tube and proceed to library quantification and quality control.

3.4 Library Quantification and Quality Control

1. This section follows protocols from Qubit, the High Sensitivity DNA Kit for the Agilent Bioanalyzer, and the Kapa Library Quantification Kit for Next Generation Sequencing.

2. Run Qubit dsDNA HS Assay as described in section 3.2 for all libraries.
3. To gain a better measure of library concentration and size distribution, run samples on Agilent 2100 Bioanalyzer HS DNA chip (*see Note 21*).
4. Up to 11 samples and 1 ladder can be run on an Agilent 2100 Bioanalyzer High Sensitivity DNA Chip. The assay requires 1 μL of sample; dilute a small volume of each original library to approximately 1 ng/ μL if necessary and run from low to high concentration in well order on the bioanalyzer chip.
5. As described in the Agilent High Sensitivity DNA Kit Guide, bring High Sensitivity DNA dye concentrate and High Sensitivity DNA gel matrix to room temperature for around 30 minutes.
6. Vortex these tubes and combine 15 μL of dye concentrate with 300 μL of gel matrix. Vortex to mix and transfer to spin filter (provided with kit).
7. Centrifuge for 10 minutes at 2300 g. Discard the filter and label the gel-dye mix with the date (it is stable when stored at 4°C in the dark for 6 weeks).
8. Ensure that the electrode is clean (*see Note 22*).
9. Place the electrode in the electrode cartridge, and insert into bioanalyzer.
10. Examine back of the HS DNA chip to ensure that it does not appear broken.
11. Pipette 9 μL of gel-dye mix into the well labeled with the G in the black circle.
12. Place chip in priming station and close lid until click.
13. Set timer for 60 seconds and steadily press plunger on syringe until it rests under the silver clip.
14. After 60 seconds, release the silver clip. Watch for the plunger to rebound to at least 0.3 mL on the syringe (more likely 0.7 mL or above) and wait for it to stop moving.

15. Remove the chip and inspect the capillaries on the back of the chip for bubbles.
16. Pipette 9 μL of gel-dye mix into the remaining 3 G wells.
17. Load 5 μL of High Sensitivity DNA marker into all remaining wells.
18. Warm samples to 37°C for 1-2 minutes (while proceeding to next step).
19. Place chip into the chip vortexer. Vortex starting from 0 rpm and ramping up to 2000-2200 rpm over a few seconds. Leave at 2200 rpm for 60 seconds.
20. Load 1 μL of sample (or water) into all sample wells and 1 μL of ladder into ladder well.
21. Vortex as above. Use chip within 5 minutes to minimize evaporation.
22. Open Expert software and select High Sensitivity DNA under Assays, DNA.
23. Once the run is complete, calculate size and concentration of libraries.
24. Following the bioanalyzer run, fragment libraries should be normally distributed with an average size of ~ 400 base pairs and jumping libraries often appear negative skewed with an average size of ~ 750 base pairs (Fig. 7). If additional peaks appear on the bioanalyzer run, these can potentially be size-selected out using an additional cleaning step with AMPure XP beads. The ratio of beads used will be based on the size of the fragment being removed and the fragment that is to be kept. [Fig 7 near here].
25. Following Qubit and Bioanalyzer quantification, libraries should undergo a final quantification step using qPCR. The KAPA kits provide standards of known size with Illumina adapters to allow the most accurate quantification of the number of adapter-ligated molecules (those that will be sequenced) in a given library.
26. Thaw all KAPA kit components on ice. Ensure that they are well mixed at each step.
27. Given that all libraries are at approximately 1-10 $\text{ng}/\mu\text{L}$, begin a serial dilution as described below (*see Note 23*).

28. Starting from the concentration described above, samples will be diluted to a final concentration of 1:200,000. This final concentration will be run alongside its 1:20,000 dilution. One possible set of volumes for this serial dilution is described here, where each successive step uses a volume from the previous step and the current concentration is listed for each step: (1) Add 1 μL library to 9 μL of water – 1:10. (2) Add 5 μL of previous dilution to 45 μL of preferred diluent – 1:100. (3) Add 5 μL of previous dilution to 45 μL of preferred diluent – 1:1000. (4) Add 5 μL of previous dilution to 45 μL of preferred diluent – 1:10,000. (5) Add 25 μL of previous dilution to 25 μL of preferred diluent – 1:20,000. (6) Add 25 μL of previous dilution to 100 μL of preferred diluent – 1:100,000. (7) Add 25 μL of previous dilution to 25 μL of preferred diluent – 1:200,000 (*see Note 24*).
29. As described in the KAPA Library Quantification Kit Technical Data Sheet, when using the kit for the first time, add 1 mL of 10X Primer Premix to 2X KAPA SYBR FAST qPCR Master Mix 5 mL bottle. Vortex reaction mixture very briefly.
30. Based on the qPCR instrument that will be used, low or high ROX might be recommended. Check manufacturer instructions.
31. Each standard (1-6), library sample (1:20,000 and 1:200,000 dilutions), and a single no template control should be run in triplicate at 20 μL per replicate.
32. Calculate the number of samples needed and create a master mix of 12 μL reaction mixture with 4 μL of water (or 3.6 μL of water with 0.4 μL of 50X ROX High or Low if necessary) for each sample (plus extra for error).
33. In qPCR plate, pipette 48 μL of master mix into each 1st replicate well.

34. Pipette 12 μ L of standard, sample, or water for no template control into each 1st replicate well and mix thoroughly.
35. Pipette 20 μ L from 1st replicate well into 2nd and 3rd replicate wells.
36. Seal plate with clear film, spin down briefly and place in instrument.
37. Set up instrument to read SYBR green and label the six 10-fold dilution standards (from 20 pM to 0.0002 pM) if possible in software.
38. Run qPCR with 95°C denaturation for 5 minutes, followed by 35 cycles of 95°C for 30 seconds and 60°C for 45 seconds with plate read (increase to 90 seconds if jumping libraries are included). Following amplification, add melt curve analysis at 65-95°C.
39. Following qPCR run, check qPCR for adapter dimers (a double peak following the main peak in the melt curve). Additional bead cleanups can remove adapter dimers and qPCR should be run again to properly quantify libraries following this step.
40. If melt curve analysis looks clean, proceed to calculation of efficiency. Remove any major outliers and plot the average C_q of the 6 standards against the log of their pM concentrations.
41. Calculate the slope and R² (which should be above 0.99). Ensure that efficiency is between 90-110% using a website such as:
<http://www.genomics.agilent.com/biocalculators/calcSlopeEfficiency.jsp>
42. If all above values are within acceptable ranges, the C_q and quantify scores given for each sample can be utilized.
43. Multiple each quantity by the original dilution (20,000 or 200,000) and size correct with the following formula: (452/library size)*average quantity. This formula corrects for the

difference in size between the library and the standards. Divide the output of this formula by 1000 to obtain nM (*see Note 25*).

44. Multiplex libraries to concentrations requested by sequencing center to fit individual budget and experimental needs.

3.5 ALLPATHS-LG Assembly

1. This section utilizes commands from the Fastq, cutadapt, TrimGalore!, and ALLPATHS-LG manuals.
2. All commands contain curly braces, { }, to indicate names (files, folders, etc.,) that need to be changed by the user to reflect the specific data being processed. Remove the curly braces and input user-specific information.

3. To run Fastqc on one fastq file, use the command:

```
fastqc --outdir {outdir_name} {fastq_file.fq}
```

4. Repeat above command line for each fastq file.
5. Run Trim Galore! on each individual library to remove adapters with commands in steps below (*see Note 26*).

6. For fragment libraries, use the command:

```
trim_galore --paired --phred33 --output_dir {outdir_name} --stringency 1 -e 0.1  
{reads1.fq} {reads2.fq}
```

7. For jumping libraries, use the command:

```
trim_galore --paired --nextera --phred33 --output_dir {outdir_name} --stringency 1 -e 0.1  
{reads1.fq} {reads2.fq}
```

8. Run Fastqc once again on trimmed reads. Verify that there are no overrepresented or adapter sequences according to the Fastqc output.

9. Write the groups and libs files required by ALLPATHS-LG. Both are comma separated files. They should be named {species}_groups.csv and {species}_libs.csv. The top line of each is the header and should appear as listed below. Asterisks are utilized in the file path to include both read directions. The library_name column must remain the same between the two files. It is underlined in this protocol.

10. Make the group csv file required by ALLPATHS-LG. This file contains 3 comma separated columns. An example is below:

group_name,library_name,file_name

{species220},{frag},{full_path_to}/{species_frag_}*fastq.gz

{species3kb},{jump3kb},{full_path_to}/{species_jump_}*fastq.gz

11. Make the libs csv file required by ALLPATHS-LG. This file contains 12 comma-separated columns. Based on the type value (either fragment or jumping), fill frag_size and frag_stddev or insert_size and insert_stddev respectively. Empty columns are allowed. An example appears below:

library_name,project_name,organism_name,type,paired,frag_size,frag_stddev,insert_size,insert_stddev,read_orientation,genomic_start,genomic_end

{frag},{avianGenome},{genusSp},fragment,1,220,{22},,,inward,,

{jump3kb},{avianGenome},{genusSp},jumping,1,,,3000,{300},outward,,

12. Create the directory structure for ALLPATHS-LG. See below:

mkdir -p {REFERENCE}/{DATA}/{RUN}/

Here, REFERENCE is the same {species} used in the groups and libs file names, DATA is arbitrary (e.g., version), and RUN is the iteration (e.g., date).

13. Prepare the ALLPATHS-LG inputs. Run command:

```
PrepareAllPathsInputs.pl DATA_DIR={full_path_to/REFERENCE/DATA/}
```

```
PLOIDY=2 IN_GROUPS_CSV=${species}_groups.csv
```

```
IN_LIBS_CSV=${species}_libs.csv > ${species}_prep_log.log
```

14. Assemble the genome. See below:

```
RunAllPathsLG PRE={full_path_to} REFERENCE_NAME={REFERENCE}
```

```
DATA_SUBDIR={DATA} RUN={RUN} HAPLOIDIFY=TRUE THREADS=
```

```
{user_choice_based_on_resources} OVERWRITE=TRUE 1> ${species}_output.log (see
```

Note 27).

15. Examine ALLPATHS-LG report (assembly.report) to view statistics.

4. Notes

1. Tissue can be fresh or frozen if, but must be of high quality for jumping library. DNA degradation can be examined on a gel as seen in Fig. 6.
2. The Apollo 324 NGS Library Preparation System was utilized alongside the Wafergen Biosystems PrepX ILM 32i DNA Library Kit for fragment library preparation in this protocol. This is not essential. The steps pertaining to use of the Apollo and this kit can be replaced with protocols from other library preparation kits if these libraries will be made entirely by hand, so long as the fragment library created this way is the appropriate size (i.e., the insert size is small enough, or the reads are long enough, that the paired-end sequencing reads will overlap).
3. When setting up the Apollo, only use Wafergen Apollo consumables unless otherwise instructed.

4. The compute required for ALLPATHS-LG is quite sizable, with our assemblies ranging from 257-389 GB of peak memory. If your lab does not have access to a supercomputer or server for this step, check with other departments at the university (e.g., physics, and structural chemistry). Two additional options available in the United States are the Amazon Web Services (available here: https://aws.amazon.com/ec2/?nc2=h_m1), and The Extreme Science and Engineering Discovery Environment (Xsede) from the National Science Foundation (available here: <https://www.xsede.org/>). To install ALLPATHS-LG, follow the link provided in the methods section to download the latest version and the ALLPATHS-LG manual. To install ALLPATHS-LG, consult the system administrator and the ALLPATHS manual. Regardless of compute resources utilized, speaking with a system or network administrator is always a good place to start when planning to install and/or run a memory-intensive program.
5. Prior to using any buffer, enzyme, etc., from a kit, ensure that it is completely thawed. Mix as directed and spin briefly (a few seconds) in a tabletop minicentrifuge to collect the volume in the bottom of the tube. Always keep enzymes on ice when not in use and return them promptly to -20°C when not in use (unless otherwise instructed).
6. Although recommended in the protocol for the E.Z.N.A. Tissue DNA Kit (as well as many other DNA extraction protocols), do not vortex your samples during DNA extraction as this could shear DNA, resulting in lower quality jumping libraries. Flick-mix by holding the tube near the lid and flicking the bottom portion repeatedly.
7. As described in the E.Z.N.A. Tissue DNA Kit Protocol, different tissues will require varied incubation times for successful lysis. Mechanically cutting the tissue into small pieces prior to the lysis will speed the process up substantially. The incubation time and

agitation method might require some optimization in your laboratory in order to get the highest quality DNA from your available tissue. Decreasing lysis time and speed of agitation could reduce shearing of DNA during lysis.

8. To obtain the most accurate quantification of samples, aim to read each tube as close to two minutes after addition of the sample/standard and the vortex mixing. Leaving approximately 30-60 seconds between mixing of each sample or standard will allow time for the reading and recording of measurements.
9. Do not allow the sample to run with visible bubbles, since these can interfere with Covaris shearing.
10. Throughout this protocol, it is recommended that Axygen Maxymum Recovery 1.7 mL microcentrifuge tubes be used to minimize product loss. Care should be taken by the unfamiliar when opening and closing these tubes, since the same properties that allow for maximum recovery of the product also provide an environment from which the product will sometimes “jump” when agitated. It is recommended that time is spent practicing with small volumes of water in these tubes before beginning a library preparation.
11. Incubation in RSB prior to elution spin with Zymo kit has been increased to 5 minutes with no discernable differences (positive or negative) on the outcome of the protocol.
12. Prior to freezing (or continuing on with the protocol), the Nextera Mate Pair Reference Guide suggests running 1 μ L of product mixed with 7 μ L of water in an Agilent DNA 12000 Bioanalyzer run to assess the tagmentation (this protocol requires a different kit than the Agilent HS DNA Bioanalyzer run described in steps 3-24 in section 3.4). Our laboratory has never carried out this quality control step for the generation of many 3 kb jumping libraries (and one 6 kb jumping library), but it can be utilized to determine the

distribution of sizes created by the tagmentation reaction in order to inform size selection ranges on the Pippin Prep. The protocol described herein appears to work well in our hands to generate 3 kb jumping libraries without the need to visualize the size distribution at this stage.

13. Add ethanol slowly to avoid disturbing beads on tube wall. If beads are disturbed, allow adequate time for beads to settle since your library is bound to the beads at this stage. Be mindful of the tendency for a sample (especially when combined with beads) to “jump” from the Axygen Maxymum Recovery 1.7 mL microcentrifuge tubes when the tubes are snapped open or closed.
14. When removing and replenishing buffer to the Elution Modules of the Pippin Prep Cassette, gently insert pipette tip into module until the bottom is reached (if the tip is not long enough or the wrong shape, use different tips). The bottom of the module is sticky or tacky. Slowly dislodge the pipette tip from the bottom and pipette as close to the bottom as possible. Avoid introducing bubbles by following this procedure and not pushing past the first resistance on the pipette (do not blow out). This same strategy applies to removing sample once the run has completed. Additionally, the starting buffer volume of 40 μ L does not fill the entire depth of the Elution Module (total space is \sim 65 μ L). Do not overfill the Elution Module.
15. There are two columns on the Pippin Prep continuity test window. If the failure appears on the left (the Separation lane), do not use the lane. If the failure appears on the right (the Elution channel), it is okay to use that channel for reference since there will be nothing to collect. Replacing the 40 μ L of electrophoresis buffer in the Elution Module

and retesting the continuity might fix a failure on the Elution lane (the right side of the window).

16. The Zymo purification step described here is a new addition to the Nextera Mate Pair Reference Guide. Our laboratory has not tested this step but based on the Nextera protocol, it should be included in here.
17. The Nextera Mate Pair Reference Guide provides an expected yield of 150-400 ng of DNA following the Pippin Prep protocol. In our laboratory, concentrations between ~1.25-5 ng/ μ L are common immediately following the Pippin Prep (prior to the new Zymo purification, while the sample is still in ~40-65 μ L). This range of concentrations (including those below the expected 150 ng total listed in the Nextera Mate Pair Reference Guide) has worked well, and there does not appear to be any discernable difference in library quality between those at the low or high end of this range. Lower yields at this step might decrease the diversity of reads, but this has not been tested in our laboratory.
18. Since bubbles can interfere with Covaris shearing, ensure that an appreciable volume of sample has not escaped during placement of the tube into the tube holder. A small bubble is likely to form at this stage (which should not be a concern), but a large bubble could be problematic. Attempt to keep the bubble contained against the cap so that results are consistent across samples.
19. Some planning is useful at this step since base diversity in the adapters is necessary for proper demultiplexing of samples after sequencing is complete. The Nextera Mate Pair Reference Guide suggests the following options if you are multiplexing 4 or fewer libraries together for a sequencing run. Use AD006 (with AD012) or AD005 (with

AD019) for 2-plexing. For 3-plexing, use AD002 (with AD007 and AD019), AD005 (with AD006 and AD015), or either of the 2-plex options with any third adapter. Use AD005 (with AD006, AD012 and AD019), AD002 (with AD004, AD007 and AD016) or any 3-plex option and any other adapter for 4-plexing. For any multiplex above 4 libraries, utilize any of the 4-plex options and then add in additional adapters as needed.

20. The number of cycles for the PCR step can be set between 10 and 15 depending on the amount of input DNA (following Pippin Prep) and concerns about library diversity. Fewer cycles will result in lower overall library yield, but more cycles can decrease library diversity by resulting in overrepresentation of highly amplified sequence. In our laboratory, 12 cycles are commonly used for most preparations with around 150-200 ng of DNA following Pippin Prep for 3 kb libraries. The Nextera Mate Pair Reference Guide suggests using only 10 cycles if there are at least 200 ng of DNA following Pippin Prep and the jump is less than 8 kb, but 15 cycles if the input DNA was lower than 200 ng regardless of insert size.
21. The Agilent Tape Station can also be used for quantification and visualization of the Apollo library, but is not compatible with the Nextera kit due to the Y-shaped adapters utilized in the jumping library.
22. A sonicator bath does an excellent job of cleaning the electrode. Following this, spray the electrode with 70% ethanol and use compressed nitrogen to dry. If a sonicator is not available, the electrode should be cleaned immediately after every run as follows: load 350 ml of nuclease-free water into Electrode Cleaning Chips and rinse the electrode by placing the chip into the bioanalyzer and closing the lid for 1 minute. Repeat 1-2 more times with fresh water. Allow electrode to air dry.

23. Depending upon the success of all library preparation steps and the number of PCR cycles that were used for each library, 1-10 ng/ μ L is a reasonable starting point. Often the jumping library will require a 1:10 dilution to reach this concentration, but if some libraries are slightly lower than this range, they could still sequence well. The KAPA kit will give the best quantification, and these values should be discussed with a technician at the sequencing center to plan your multiplex.
24. Possibly the most important aspect of obtaining accurate and precise measures from the KAPA qPCR step is effective sample mixing. Small pipetting errors will accumulate quickly, and if dilutions are not homogeneous at the time of pipetting, this will also add error. Pipette-mixing 20-50 times at each mixing step appears to reduce noise in the final data.
25. Following KAPA library quantification, the technicians at our sequencing center recommend correcting the final nM concentration by a factor of 0.75 before multiplexing (e.g., library at 1 nM is reported as 0.75 nM). In their experience, this factor produces optimal cluster density on their sequencers. Contact the technicians at your sequencing center to determine if they have recommendations for any correction following KAPA quantification.
26. Although we currently recommend Trim Galore! as written in the methods section for the adapter trimming, all genomes discussed in this chapter were trimmed using Trimmomatic. It requires the generation of an adapters file that provides Trimmomatic with the specific adapter sequences utilized (e.g., Nextera and Apollo). The command for this is available here: github.com/tsackton/ratite-genomics/blob/master/assembly/trim_fastq.sh

27. THREADS= {user_choice_based_on_resources} is the number of threads requested based on system infrastructure. Additionally, if assembling an inbred or low polymorphism genome, HAPLODIFY=FALSE might provide better results, but our genomes from wild and zoo specimens have been assembled with better success using HAPLODIFY=TRUE.

5. References

1. C.D. Stern (2005) The chick: A great model system becomes even greater, *Developmental Cell*. 8, 9–17.
2. H. Nagai, S.-S. Mak, W. Weng, et al. (2011) Embryonic development of the emu, *Dromaius novaehollandiae*., *Developmental dynamics : an official publication of the American Association of Anatomists*. 240, 162–75.
3. C.S. Padgett and W.D. Ivey (1960) The normal embryology of the coturnix quail, *The Anatomical Record*. 137, 1–11.
4. J.R. Murray, C.W. Varian-Ramos, Z.S. Welch, et al. (2013) Embryological staging of the Zebra Finch, *Taeniopygia guttata*, *Journal of Morphology*. 274, 1090–1110.
5. L.W. Hillier, W. Miller, E. Birney, et al. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution., *Nature*. 432, 695–716.
6. G. Zhang, C. Li, Q. Li, et al. (2014) Comparative genomics reveals insights into avian genome evolution and adaptation, *Science*. 346, 1311–1320.
7. E.D. Jarvis, S. Mirarab, A.J. Aberer, et al. (2014) Whole Genome Analyses Resolve the Early Branches in the Tree of Life of Modern Birds, *Science*. 346, 1320–1331.
8. G. Zhang (2015) Bird sequencing project takes off, *Nature*. 522, 34.
9. C. Bonneaud, J. Burnside, and S. V Edwards (2008) High-Speed Developments in Avian Genomics, 58, 587–595.
10. S. Gnerre, I. Maccallum, D. Przybylski, et al. (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data., *Proceedings of the National Academy of Sciences of the United States of America*. 108, 1513–1518.
11. K.R. Bradnam, J.N. Fass, A. Alexandrov, et al. (2013) Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species., *GigaScience*. 2, 10.
12. B. Langmead and S.L. Salzberg (2012) Fast gapped-read alignment with Bowtie 2, *Nat Methods*. 9, 357–359.
13. S. Goodwin, J. Gurtowski, S. Ethe-Sayers, et al. (2015) Oxford Nanopore Sequencing and de novo Assembly of a Eukaryotic Genome, *bioRxiv*. 13490.
14. A. Rhoads and K.F. Au (2015) PacBio Sequencing and Its Applications, *Genomics, Proteomics and Bioinformatics*. 13, 278–289.
15. D. Gordon, J. Huddleston, M.J. Chaisson, et al. (2016) Long-read sequence assembly of the gorilla genome, *Science*. 352, aae0344.
16. Y. Mostovoy, M. Levy-Sakin, J. Lam, et al. (2016) A hybrid approach for de novo human genome sequence assembly and phasing., *Nature methods*. 13, 12–17.
17. N.I. Weisenfeld, V. Kumar, P. Shah, et al. (2016) Direct determination of diploid genome sequences, *bioRxiv*. 70425.
18. A. Siepel, G. Bejerano, J.S. Pedersen, et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes, *Genome Research*. 15, 1034–1050.
19. A. Visel, S. Prabhakar, J.A. Akiyama, et al. (2008) Ultraconservation identifies a small subset of extremely constrained developmental enhancers, *Nat Genet*. 40, 158–160.
20. C.B. Lowe, M. Kellis, A. Siepel, et al. (2011) Three periods of regulatory innovation during vertebrate evolution., *Science (New York, N.Y.)*. 333, 1019–1024.
21. C.B. Lowe, J.A. Clarke, A.J. Baker, et al. (2014) Feather Development Genes and Associated Regulatory Innovation Predate the Origin of Dinosauria, 32, 23–28.
22. A. Marcovitz, R. Jia, and G. Bejerano (2016) “Reverse Genomics” Predicts Function of

- Human Conserved Noncoding Elements, *Molecular Biology and Evolution*. 33, 1358–1369.
23. M. Hiller, B.T. Schaar, and G. Bejerano (2012) Hundreds of conserved non-coding genomic regions are independently lost in mammals, *Nucleic Acids Research*. 40, 11463–11476.
 24. B.M. Booker, T. Friedrich, M.K. Mason, et al. (2015) Bat Accelerated Regions Identify a Bat Forelimb Specific Enhancer in the *HoxD*,.
 25. W.L. Eckalbar, S.A. Schlebusch, M.K. Mason, et al. (2016) Transcriptomic and epigenomic characterization of the developing bat wing, *Nature Genetics*. 48, 528–536.
 26. E.T. Domyan, Z. Kronenberg, C.R. Infante, et al. (2016) Molecular shifts in limb identity underlie development of feathered feet in two domestic avian species, *eLife*. 5, 1–21.
 27. N. Adachi, M. Robinson, A. Goolsbee, et al. (2016) Regulatory evolution of *Tbx5* and the origin of paired appendages, *Proceedings of the National Academy of Sciences*. 113, 10115–10120.

Table and Figure Titles

Figure 1. Contig N50 in kilobases versus total number of sequencing reads (3 kb jumping and fragment libraries combined) for 16 draft avian genomes assembled using ALLPATHS-LG.

Figure 2. Scaffold N50 in kilobases versus total number of sequencing reads (3 kb jumping and fragment libraries combined) for 16 draft avian genomes assembled using ALLPATHS-LG.

Figure 3. Contig N50 in kilobases versus scaffold N50 in megabases for 17 draft avian genomes assembled in our laboratory alongside published avian genomes available from NCBI categorized by assembly method (total n=72). NCBI avian genomes were included if the metadata contained assembly method, contig N50 and scaffold N50 values. This figure does not display the chicken or great tit (*Parus major*) since their contig N50 (chicken) and scaffold N50 (great tit) values are 5 and 4 times greater than all the other species respectively. ALLPATHS-LG provides excellent contig N50 values and good scaffold N50 values across avian samples. Since gene annotation (and mapping of short reads for epigenomic assays) relies more heavily on large contigs than large scaffolds, the ALLPATHS-LG approach is appropriate for developmental genomics.

Figure 4. A draft genome provides a developmental research group with the opportunity to explore many areas, including gene expression, gene manipulation, gene regulation and genome evolution.

Figure 5. Overview of methods presented in this chapter to produce ALLPATHS-LG genome assembly starting from high quality avian tissue.

Figure 6. Composite image of a 1kb ladder (1 kb) and three DNA extractions using slightly different methods on tissues stored under identical conditions (frozen in RNAlater at -80°C) to show differences in DNA quality. The first sample (MR) was muscle tissue that was lysed for only 3 hours and treated with RNase A before extraction, the second (LR) is liver tissue treated in the same way, and the last sample (LNR – far right) is a piece from the same liver sample after being lysed overnight (18 hours) without the addition of RNase A. RNase A (or the shorter lysis step) appears to have cleaned up some of the extended smear seen in LNR compared to LR. We utilized MR for genome sequencing since it appeared to have the least degradation.

Figure 7. Example electropherograms from an HS DNA Assay run on the Agilent 2100 Bioanalyzer for a Nextera 3 kb jumping library (A) and Apollo 220 bp fragment library (B). In the Gel image, the library is present between the Lower Marker (LM) at the bottom of the image, and the Upper Marker (UM) at the top. Average sizes are approximately 750 and 400 bp for the jumping and fragment libraries respectively.

Table 1. Assembly statistics for 17 unpublished avian genomes assembled using ALLPATHS-LG.

[Table 1]

Common Species Name	Latin Species Name	Estimated Genome Size (Gb)	Total Coverage	Contig N50 (Kb)	Scaffold N50 (Mb)	Fragment Reads (E+08)	Jump Reads (E+08)	Total Reads (E+08)
Great spotted kiwi	<i>Apteryx haastii</i>	1.35	44.4	106	1.36	2.95	2.77	5.72
Little spotted kiwi	<i>Apteryx owenii</i>	1.36	60.5	128	1.62	4.48	4.52	8.99
Okarito brown kiwi	<i>Apteryx rowi</i>	1.40	54.3	120	1.67	3.58	4.36	7.94
Black-footed albatross	<i>Phoebastria nigripes</i>	1.25	67.1	81.9	1.53	4.40	3.57	7.98
Northern cardinal	<i>Cardinalis cardinalis</i>	1.10	58.8	114	3.66	3.36	3.68	7.04
Southern cassowary	<i>Casuarus casuarus</i>	1.37	54.1	133	3.70	4.19	4.6	8.79
Thicket tinamou	<i>Crypturellus cinnamomeus</i>	1.2	47.5	50.3	2.43	2.70	3.22	5.92
Emu	<i>Dromaius novaehollandiae</i>	1.31	64.8	139	3.32	5.01	3.98	8.99
Elegant crested tinamou	<i>Eudromia elegans</i>	9.64	68.0	98.8	3.28	3.43	3.29	6.72
Laysan albatross	<i>Phoebastria immutabilis</i>	1.25	59.4	81.4	1.62	3.97	2.96	6.93
Leach's storm petrel	<i>Oceanodroma leucorhoa</i>	1.24	79.5	165	8.58	4.40	5.83	10.2
Chilean tinamou	<i>Nothoprocta perdicaria</i>	1.01	76.2	75.8	3.35	4.29	3.36	7.66
Greater rhea	<i>Rhea americana</i>	1.24	59.9	68.7	4.08	3.90	3.76	7.66
Lesser rhea	<i>Rhea pennata</i>	1.27	42.4	55.9	3.85	3.13	3.31	6.44
Short-tailed albatross	<i>Phoebastria albatrus</i>	1.23	41.5	56.7	1.52	2.27	2.65	4.92
Waved albatross	<i>Phoebastria irrorata</i>	1.24	41.1	63.0	1.51	2.24	2.76	5.00
Wandering albatross	<i>Diomedea exulans</i>	1.22	28.3	42.8	1.10	1.50	2.06	3.56

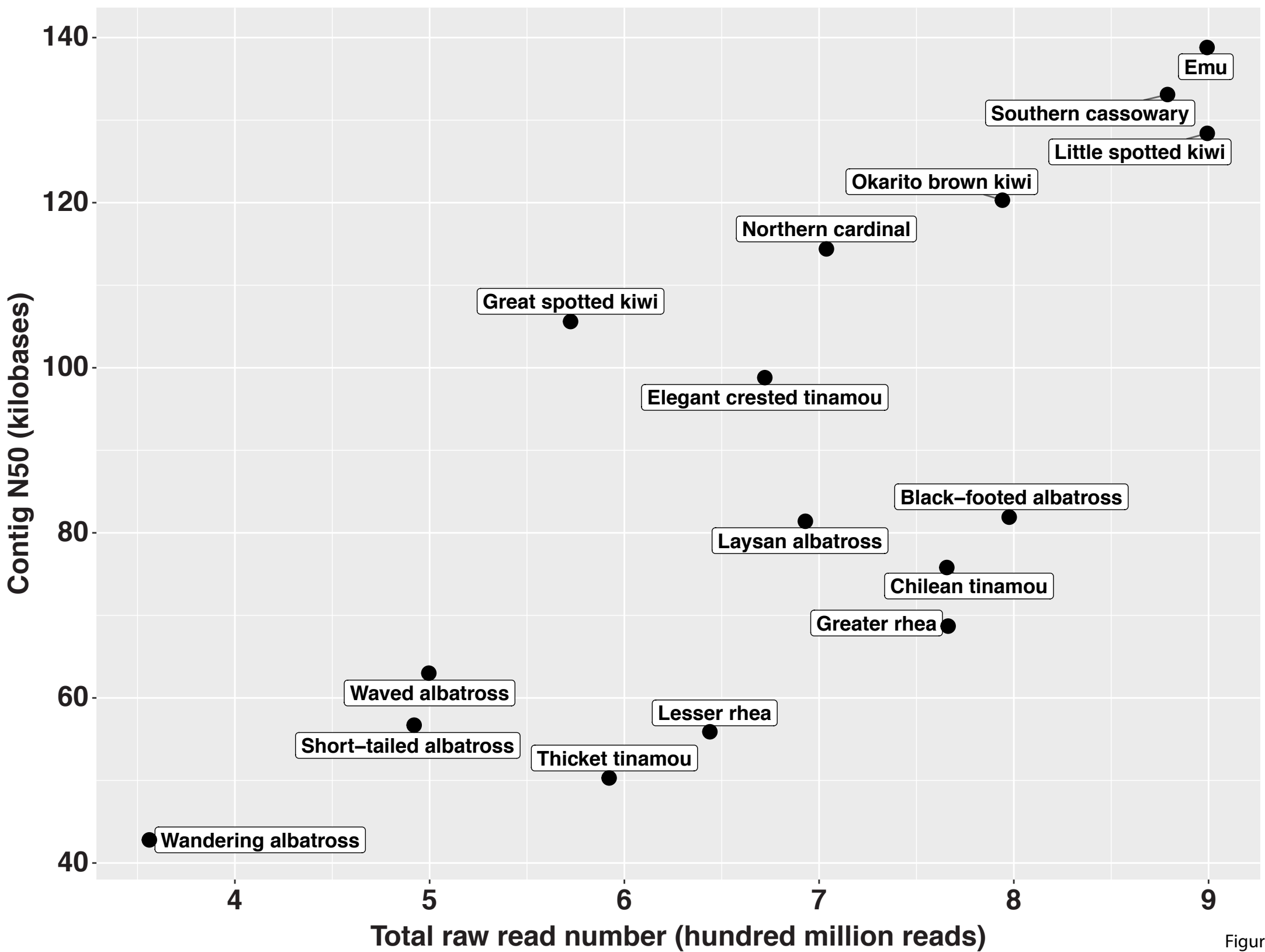


Figure 1

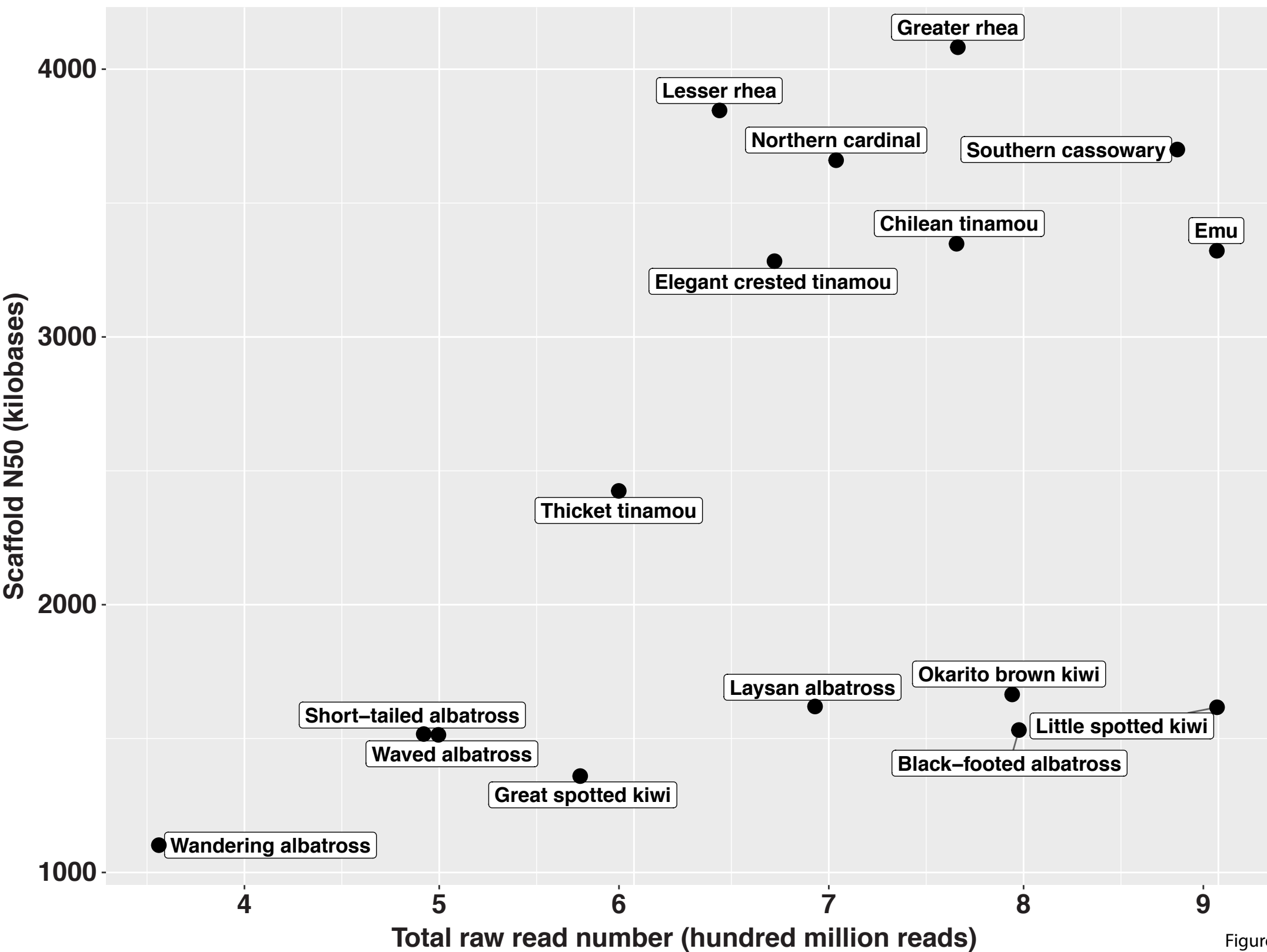


Figure 2

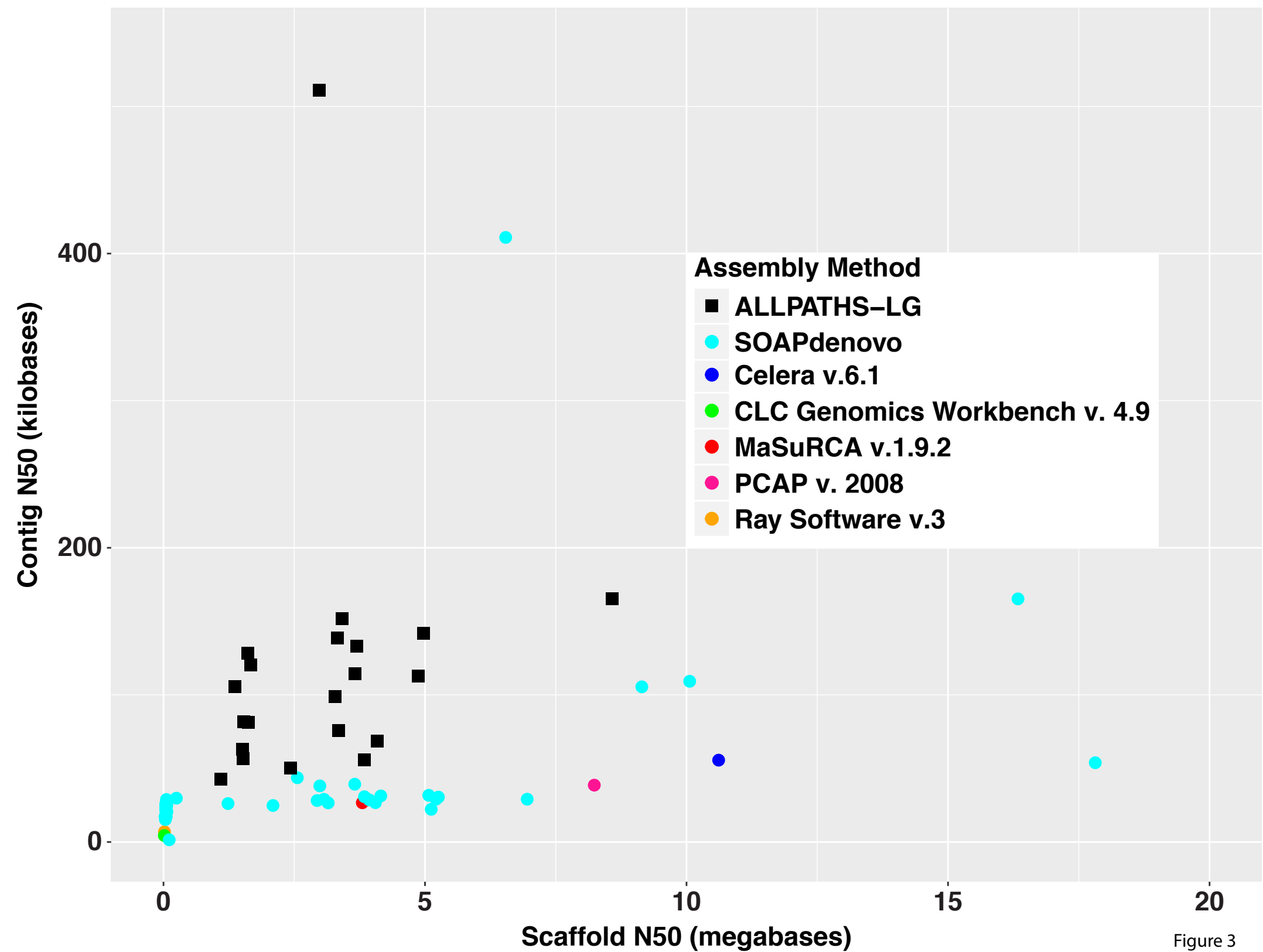
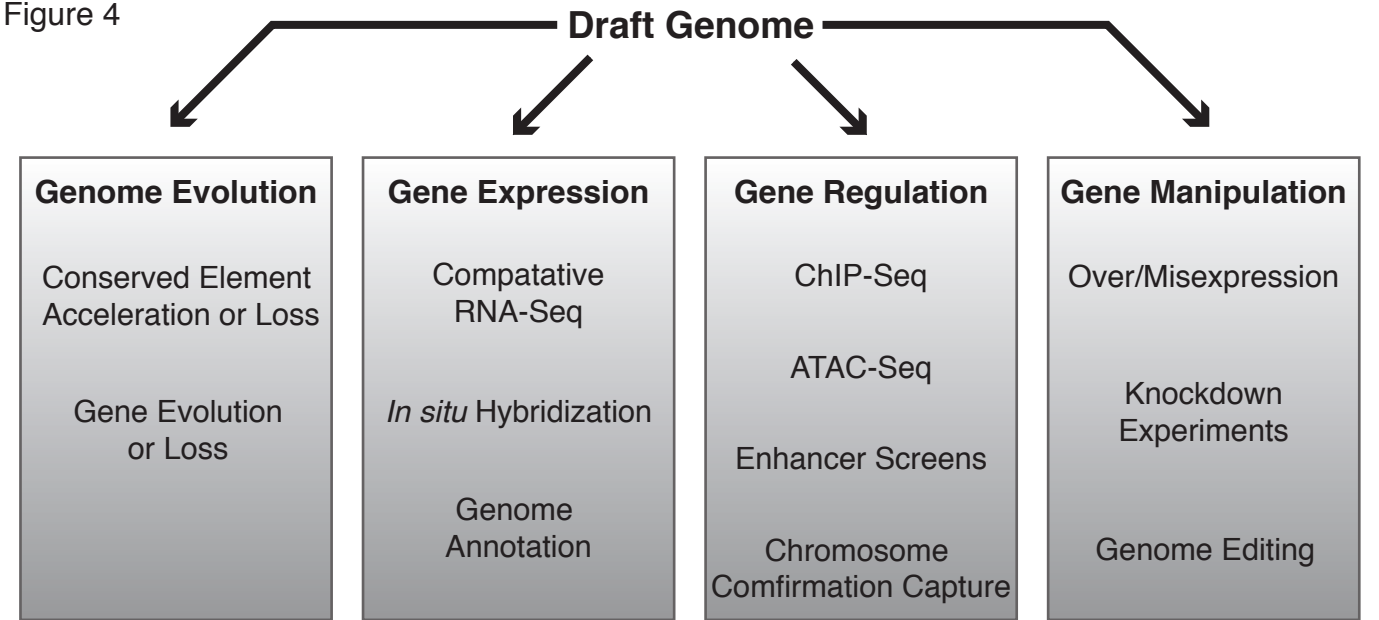


Figure 3



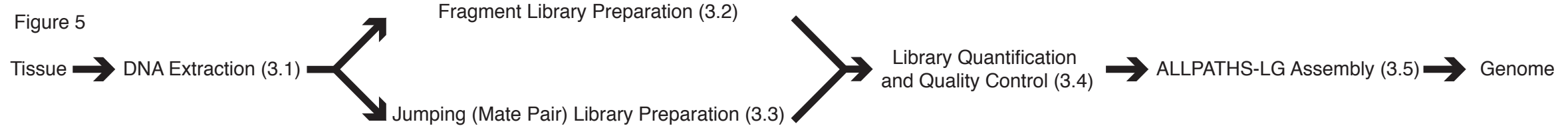


Figure 6

1 kb

MR

LR

LNR

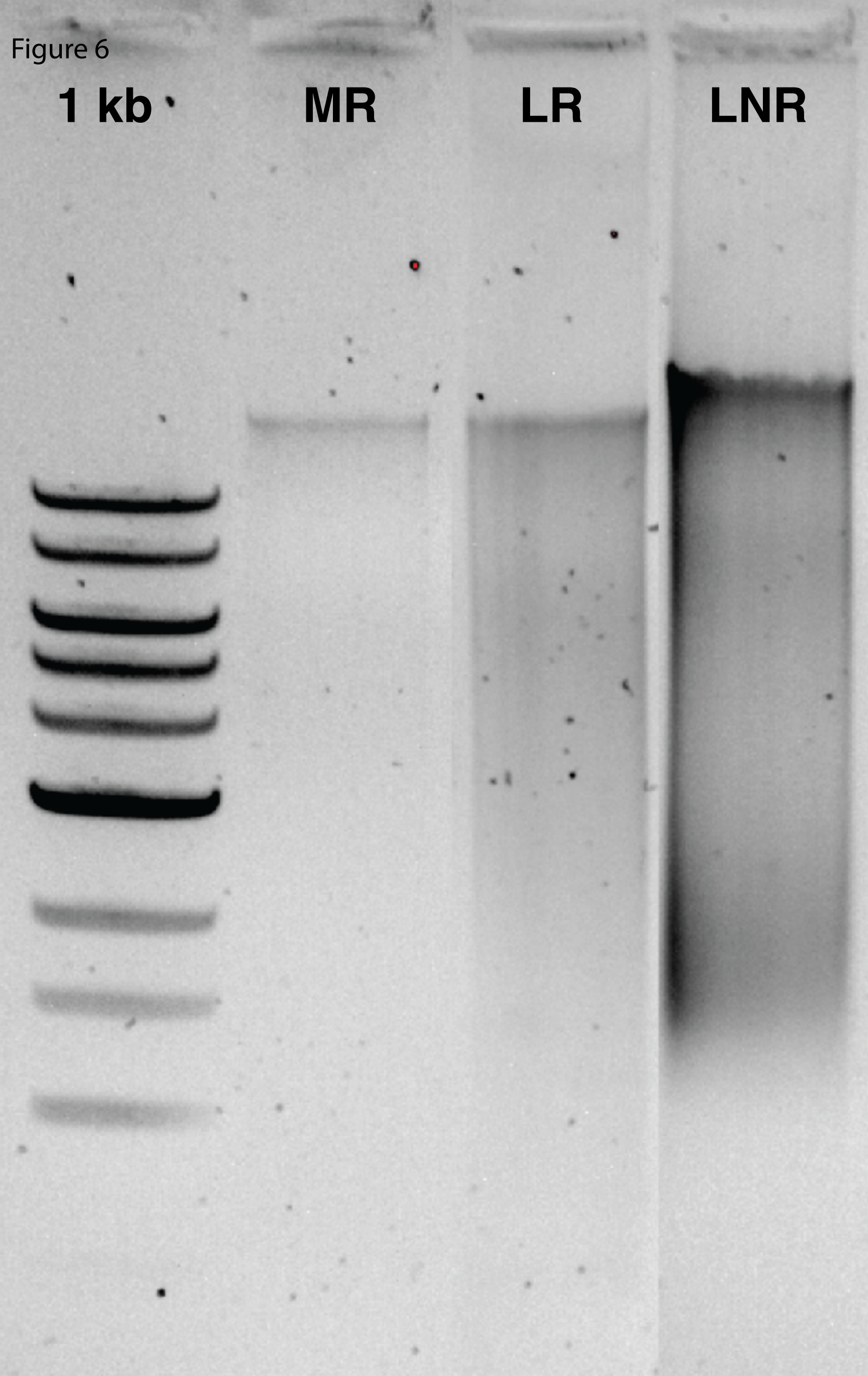


Figure 7

