Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext)

Travis C. Glenn[1-5,*], Roger A. Nilsen[4], Troy J. Kieran[1], John W. Finger Jr.[1,2], Todd W. Pierson[1,§], Kerin E. Bentley[3], Sandra L. Hoffberg[3], Swarnali Louha[5], Francisco J. García-De León[6], Miguel Angel Del Rio Portilla[7], Kurt D. Reed[8], Jennifer L. Anderson[9], Jennifer K. Meece[9], Samuel E. Aggrey[5,10], Romdhane Rekaya[5,11], Magdy Alabady[4,12], Myriam Bélanger[4,13], Kevin Winker[14], and Brant C. Faircloth[15,*]

[1]Department of Environmental Health Science, University of Georgia, Athens, GA 30602, USA
[2]Interdisciplinary Toxicology Program, University of Georgia, Athens, GA 30602, USA
[3]Department of Genetics, University of Georgia, Athens, GA 30602, USA
[4]Georgia Genomics Facility, University of Georgia, Athens, GA 30602, USA
[5]Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA
[6]Laboratorio de Genética para la Conservación, Centro de Investigaciones Biológicas del Noroeste, SC, Instituto Politécnico Nacional 195, Playa Palo de Santa Rita Sur, La Paz, B.C.S., C.P. 23096, México
[7]Departamento de Acuicultura, CICESE. Carretera Ensenada-Tijuana 3918 Zona Playitas, Ensenada Baja California, 22860, México
[8]Department of Pathology and Laboratory Medicine, University of Wisconsin, Madison, WI 53792, USA
[9]Integrated Research and Development Laboratory, Marshfield Clinic Research Foundation, Marshfield, WI 54449, USA
[10]Department of Poultry Science, University of Georgia, Athens, GA 30602, USA
[11]Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA
[12]Department of Plant Biology, University of Georgia, Athens, GA 30602, USA
[13]Department of Infectious Diseases, University of Georgia, Athens, GA 30602, USA
[14]University of Alaska Museum, 907 Yukon Drive, Fairbanks, AK 99775, USA
[13]Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA

[§]current address: Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA

Keywords: combinatorial, DNA libraries, Illumina, iNext, iTru, multiplexing, Nextera, NGS, TruSeq

[*]**Correspondence:** Travis C. Glenn, Dept. of EHS, Environmental Health Science Bldg., University of Georgia, Athens, GA 30602, USA; Tel: 706-583-0662; fax: 706-542-7472; E-mail: travisg@uga.edu

Brant C. Faircloth, Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA; Tel: 225-578-1006; fax: 225-579-2597; E-mail: brant@faircloth-lab.org

Running Title: Universal Adapter System for Illumina Libraries

ABSTRACT

Next-generation DNA sequencing (NGS) offers many benefits, but major factors limiting NGS include reducing the time and costs associated with: 1) start-up (i.e., doing NGS for the first time), 2) buy-in (i.e., getting any data from a run), and 3) sample preparation. Although many researchers have focused on reducing sample preparation costs, few have addressed the first two problems. Here, we present iTru and iNext, dual-indexing systems for Illumina libraries that help address all three of these issues. By breaking the library construction process into re-usable, combinatorial components, we achieve low start-up, buy-in, and per-sample costs, while simultaneously increasing the number of samples that can be combined within a single run. We accomplish this by extending the Illumina TruSeq dual-indexing approach from 20 (8+12) indexed adapters that produce 96 (8x12) unique combinations to 579 (192+387) indexed primers that produce 74,304 (192x387) unique combinations. We synthesized 208 of these indexed

primers for validation, and 206 of them passed our validation criteria (99% success). We also used the indexed primers to create hundreds of libraries in a variety of scenarios. Our approach reduces start-up and per-sample costs by requiring only one universal adapter which works with indexed PCR primers to uniquely identify samples. Our approach reduces buy-in costs because: 1) relatively few oligonucleotides are needed to produce a large number of indexed libraries; and 2) the large number of possible primers allows researchers to use unique primer sets for different projects, which facilitates pooling of samples during sequencing. Although the methods we present are highly customizable, resulting libraries can be used with the standard Illumina sequencing primers and demultiplexed with the standard Illumina software packages, thereby minimizing instrument and software customization headaches. In subsequent *Adapterama* papers, we use these same iTru primers with different adapter stubs to construct double- to quadruple-indexed amplicon libraries and double-digest restriction-site associated DNA (RAD) libraries. For additional details and updates, please see http://baddna.org.

Introduction

Next-generation sequencing (NGS) is transforming the life sciences because the unprecedented amount of sequence data generated by NGS platforms facilitates new approaches, techniques, and discoveries (Ansorge 2009, Tautz *et al*. 2010). Reduced costs (Glenn 2011, 2016) are a major component of NGS success because cost reduction enables many studies that were previously infeasible. Although NGS costs per read have dropped tremendously, the minimum cost to obtain any amount of NGS data (i.e., the minimum buy-in cost) remains high, particularly when researchers want to collect small amounts of DNA sequence data from large numbers of individual samples in a single run. These buy-in costs are largely driven by the money required

to purchase adapters containing unique identifying sequences that allow tagging and tracking of samples sequenced in multiplex (Box 1). For example, the purchase price for a subset of 96, single-index, TruSeq-equivalent adapters described in Faircloth and Glenn (2012) would require an initial investment of $4,019 (US), and this investment is exclusive of the additional costs to purchase other necessary library preparation reagents and consumables. A second problem for researchers wishing to collect smaller amounts of sequence data from many samples sequenced in multiplex is the relatively limited number of indexed adapters that are available. Although several publications (e.g., Meyer & Kircher 2010, Faircloth & Glenn 2012, Rohland & Reich 2012) and commercial products (e.g., Illumina Nextera, Bioo Scientific NEXTflex-HT) provide schemes for indexing hundreds of individuals sequenced in multiplex, these approaches do not facilitate individually tagging many thousands of samples at low cost so that many possible subsets of hundreds of samples can be pooled into a single sequencing run. As a result, library preparation methods that reduce costs while simultaneously increasing the number of samples that can be tagged and sequenced together would benefit many types of research.

In this first paper of the *Adapterama* series, we present the key components of an integrated system for producing large numbers (>70,000) of uniquely tagged, dual-indexed Illumina libraries at low cost (Figs. 1, S1). We build this integrated system on top of previous developments introduced by Illumina (2008) and others (e.g., Meyer and Kircher 2010, Fisher *et al*. 2011), and we show that it is possible to significantly reduce library preparation costs by changing from full-length adapters that incorporate tags in the Illumina TruSeq strategy to shorter universal adapter stubs and indexing primers (hereafter referred to as the iTru strategy; which is similar to the original Illumina indexing strategy [Illumina 2008]). Simply moving from a TruSeq indexing strategy to the iTru indexing strategy, while maintaining a single

105 indexing position, can reduce costs by more than 50% (Table 1). When taking advantage of the
106 dual-indexing offered by our iTru strategy, researchers can reduce costs by at least an order of
107 magnitude relative to TruSeq (Table 1). This method is also extensible to the Illumina Nextera
108 adapter sequences (Syed *et al.* 2009, Adey *et al.* 2010), hereafter referred to as the iNext
109 approach (Figs. S1, S2; Supplemental File 1). We focus on describing the iTru system because
110 TruSeq is more commonly used than Nextera and to simplify presentation of the system (details
111 of the iNext system are generally given in the supplemental figures and files). In subsequent
112 *Adapterama* manuscripts, we modify the system presented here for a variety of applications (e.g.,
113 amplicon sequencing, RAD-sequencing), but we use our iTru or iNext indexing primers
114 throughout (Fig. S1).
115 Here we outline the ideas underlying genomic library construction for Illumina sequencers,
116 and we provide some historical perspective on Illumina library preparation for researchers new
117 to Illumina sequencing. Following this introduction, we describe our iTru approach, which
118 modifies Illumina's original library construction method and extends the approach to include
119 indexes on both primers (i.e., double-indexing; c.f., Kircher *et al*. 2012). The iTru approach
120 (Figs. 1-3) produces: (1) libraries that are compatible with all Illumina sequencing instruments
121 and reagents, (2) libraries that can be pooled (i.e., multiplexed) with other Illumina libraries, (3)
122 libraries that can be sequenced using standard Illumina sequencing primers and protocols, and
123 (4) data that can be demultiplexed with standard Illumina software packages and pipelines.
124
125 *Illumina Libraries*
126 DNA molecules that can be sequenced on Illumina instruments require specific primer-
127 binding sites (i.e., adapters; Box 1) on each end. The procedure to incorporate the adapters to the

DNA insert is generally referred to as "library preparation". Library preparation of genomic DNA, in its most common form, involves randomly shearing DNA to a desired size range (e.g., 200-600 bp); end-repairing and adenylating the sheared DNA; adding synthetic, double-stranded adapters onto each end of the adenylated DNA molecules using T/A ligation; and using limited-cycle PCR amplification to increase the copy number of valid constructs (Figs. 1-3, S3; cf. S2, S4).

Illumina library preparations differed from their early competitors (chiefly 454) because their double-stranded adapters used a Y-yoke design to increase library construction efficiency (Bentley *et al*. 2008; Greigite 2009). The Y-yoke structure of the adapters allows each starting DNA molecule to serve as two templates, requiring ≥3 cycles of PCR to produce complete double-stranded library molecules (Fig. S3). The DNA molecules resulting from these preparations (Figs. 1, 3, S4) contain (a) outer primer-binding sites (P5 and P7) used to capture individual DNA molecules on the surface of Illumina flow cells and clonally amplify them, (b) separate primer-binding sites (Read 1 and Read 2), located internal to the P5 and P7 sites, that allow directional sequencing of both DNA strands, and (c) short DNA sequences, known as indexes (Box 1; see below), inserted into the P7 side of the adapter molecule (Illumina, 2008; Fig. 4 - i7 index, sequence obtained from Index Read 1).

Indexing

Indexing strategies are generally meant to individually identify different DNA samples by incorporating unique DNA sequences into the library constructs (Shoemaker *et al.* 1996, Binladen *et al*. 2007, Glenn *et al*. 2007, Hoffmann *et al*. 2007, Meyer *et al*. 2007, Craig *et al.* 2008). Indexed libraries can then be pooled together (multiplexed) in a single sequencing lane.

151 During sequencing, individual molecules are captured on the surface of the Illumina flow cells,
152 the individual molecules are clonally amplified, and up to four separate sequencing reactions
153 take place sequentially, each creating a separate sequencing read (Fig. 4). After sequencing,
154 computer software matches the observed index sequence for each molecule to a list of samples
155 with expected indexes (i.e., using a sample sheet; Supplemental File 2) and parses the bulk data
156 back into its component parts (i.e., demultiplexed, e.g., using bcl2fastq [Illumina 2013]).
157 In practice, the history and current status of Illumina indexing strategies is quite complicated
158 (e.g., Illumina 2016a), with several transitions among different adapter systems that resulted
159 from changing capabilities of sequencing instruments. Illumina originally created 12 different i7
160 indexes (Figs. 1, 3, 4) to allow pooling of up to 12 samples, and the company later increased the
161 number of i7 indexes for certain applications to 48. The original Illumina i7 indexes had a length
162 of six nucleotides (nt) and were constructed such that ≥2 substitution errors were needed to turn
163 one index into another – an effort to minimize sample confusion as a result of sequencing error.
164 Sequencing errors on Illumina instruments are primarily substitutions, thus Illumina's initial
165 indexes were designed to be robust to substitution sequencing errors. Deletions, however, are
166 the primary errors of oligonucleotide synthesis (i.e., synthesis of the adapters and/or primers used
167 to make the indexed libraries). It is, therefore, desirable to have indexes that are robust to
168 insertions and deletions (indels) as well as substitutions, thus conforming to an edit-distance
169 metric and limiting the assignment of sequences to the wrong sample (Faircloth and Glenn
170 2012). When index sets have distances ≥3, then error correction can be employed, but this
171 distance criterion is frequently violated (Faircloth and Glenn 2012).
172 Building upon earlier in-house and external efforts, Illumina introduced a product (Nextera
173 kits) that used an i5 index and an i7 index (i.e., dual-indexing; see Box 1, Fig. 1, and below) each

of which were longer (8 nt) and, at that time, conformed to the edit-distance metric. Nextera adapters use the same sequences for interaction with the flow-cell (i.e., P5 and P7, Fig. 1), but have unique Read 1 and Read 2 sequences relative to TruSeq (Figs S2, S4), and thus Illumina does not recommend combining Nextera and TruSeq libraries within a single sequencing lane (Illumina 2012; but see below). Illumina subsequently incorporated 8 nt, dual indexes into the TruSeq system with their release of TruSeqHT. Although the Illumina TruSeqHT indexes are robust to insertion, deletion, and substitution errors, the updated TruSeqHT i7 indexes do not maintain an edit distance ≥3, when compared to other TruSeq HT i7 indexes in the same set or when combined with all previous Illumina i7 indexes, and so do not allow proper error correction (Figure S5; Supplemental File 3). Regardless, the TruSeqHT indexing system is more robust, accurate, and flexible than previous approaches, and researchers can index template DNA molecules using the i7 indexes alone (single indexing) or in combination with i5 indexes (dual indexing).

Dual indexing on the Illumina platform means that tags can be used combinatorially (Kircher *et al*. 2012; Faircloth and Glenn 2012). Major advantages of the dual indexing strategy include: a) the need for fewer oligonucleotides to index the same number of samples in multiplex (e.g., 8 + 12 = 20 primers produce 8 x 12 = 96 unique tag combinations), b) concomitantly reducing the cost of production, inventory, and quality control (QC) (i.e., it is less expensive to produce, maintain stocks of, and do QC on 20 primers than 96), and c) the universality of the approach – dual-indexing is compatible with both full-length adapters (e.g., TruSeqHT libraries) or universal adapter stubs and primers (e.g., Nextera, iNext, or iTru).

Illumina-compatible Libraries

Illumina's libraries have been the industry's gold standard for sequence quality on Illumina platforms, but their library preparation kits are among the most expensive available. The number of indexes offered by Illumina has been limited to ≤48 and the number of dual-index combinations ≤96, until the relatively recent release of additional indexes for the Nextera system, which can double-index up to 384 samples. Alternative commercial kits have been produced to increase efficiency, reduce GC bias (Aird *et al.* 2011, Kozerewa *et al.* 2009), and/or increase the number of indexes, but costs remain high and the total number of commercially available tags still generally remains ≤384.

A variety of library preparation methods have also been described by research groups that reduce per-sample costs relative to most commercial kits (e.g., Meyer and Kircher 2010 [MK-2010], Fisher *et al*. 2011 [F-2011]; see Head *et al*. 2014 for others). The MK-2010 and F-2011 methods are in widespread use, but they do have some shortcomings. For example, the MK-2010 method: a) specifies HPLC purification of adapter oligonucleotides, which increases start-up costs dramatically and can lead to contamination from previous oligonucleotides that were purified on the same HPLC columns; b) relies on hairpin suppression of molecules with identical adapter ends (instead of using a Y-yoke adapter) which is efficient with smaller inserts (e.g., <200 bp) but loses efficiency with increasing insert length; and c) relies on blunt-ended ligation, which allows the formation of chimeric inserts, a danger that increases with insert length. The F-2011 method introduced the idea of "on-bead" library preparation, which increases efficiency and reduces costs; thus, many commercial kits have subsequently incorporated similar on-bead library preparation approaches. Limitations of the F-2011 method include use of: a) custom NEB reagents, not in the standard catalog or available in small quantities; b) large volumes of

enzymes; and c) Illumina adapters and primers, which increase costs and limit the number of samples that can be pooled.

Our approach builds upon many of the previous approaches introduced by Illumina, MK-2010, F-2011, Rohland & Reich (2012), and others to develop library preparation methods for genomic DNA that overcome many of these limitations. We describe adapters, primers, and library construction methods that produce DNA molecules equivalent to and compatible with Illumina's TruSeqHT libraries (and, separately, Nextera libraries, see Supplemental File 1; Table 2). Our method extends the number of available index combinations from 8 x 12 to 192 x 387, while maintaining a minimum edit-distance of three between all indexes. We demonstrate the effectiveness of our combinatorial indexing primers by controlled quantitative PCR experiments, and we demonstrate the utility of our system by preparing and sequencing iTru libraries from organisms with varying genome size and DNA quality.

Methods

Adapter and primer design

We modified the Illumina TruSeq system by dividing the adapter components into two parts: 1) a universal Y-yoke adapter "stub" that comprises parts of the Read 1 and Read 2 primer binding sites plus the Y-yoke, and 2) a set of amplification primers (iTru5, iTru7), parts of which are complementary to the Y-yoke stub and which also contain custom sequence tag(s) for sample indexing (Figs. 1, 3; Table 3; Supplemental File 4). The iTru Y-yoke adapter has a single 5' thymidine (T) overhang and can be used in standard library preparations that produce insert DNA with single 3' adenosine (A) overhangs. We designed a large set of indexed amplification primers (iTru5, iTru7; Supplemental File 4) that contain a subset of our custom 8 nt sequence

242 tags (from Faircloth & Glenn 2012), as well as an initial set that incorporated the TruSeq HT
243 indexes (i.e., D5xx for iTru5 and D7xx for iTru7) which could serve as controls. We grouped
244 the iTru primers with our sequence tags into clearly identifiable, numbered sets (100 or 300
245 series) that are compatible with 8 nt tags in the standard Illumina TruSeqHT primers, as well as
246 Illumina v2 8 nt tags (including the 6 nt tags converted to 8 nt via addition of invariant bases
247 from the adapter). We also created several additional numbered sets (200 or 400 series) of iTru
248 primers that are compatible with all other primers and sequence tags in our iTru system, but
249 which are not compatible with all Illumina indexes. We then balanced the base composition of
250 all iTru primers in all numbered sets in groups of eight for iTru5 or 12 for iTru7, because
251 balanced base composition is critical for successful index sequencing (Illumina 2016b; see
252 Discussion for additional information on combining small numbers of libraries).
253 We ordered the components of our Y-yoke adapter stubs and iTru primers from
254 Integrated DNA Technologies (IDT, Coralville, IA, USA). We modified the adapter stub
255 sequence by phosphorylating the 5' end of iTru_R2_stub_RCp oligonucleotide (Figure 1; Table
256 3), and we modified each of the iTru primer sequences by adding a phosphorothioate bond
257 (Eckstein 1985) before the 3' nucleotide of each sequence to inhibit degradation due to the
258 exonuclease activity of proof-reading polymerases (Skerra 1992), which are commonly used in
259 library preparation. Following initial small-scale orders, we ordered the entire complement of
260 iTru primers, placing the iTru5 and iTru7 primers into every other column (iTru5) or row (iTru7)
261 of 96-well plates, with 0.625 or 1.25 nmol aliquots in replicate plates (Supplemental Files 4, 5).
262 We hydrated newly synthesized primers to 10 µM in the plate and 5 µM prior to use
263 (Supplemental File 6).
264

Validation of iTru Primers by Quantitative PCR (qPCR)

To determine whether our indexed iTru5 and iTru7 primers were biasing amplification, we selected a subset of iTru7 (n=160) and iTru5 (n=48) primers for qPCR validation. To validate the iTru primers, we prepared a pool of adapter-ligated chicken DNA using an inexpensive, double-digest RAD approach (3RAD; Glenn *et al.* 2016b) that produces a DNA construct having 5' and 3' ends identical to our Y-yoke adapter. We then set up quantitative PCR reactions with 5 µL GoTaq qPCR Master Mix (Promega, Madison, WI, USA), 1 µL each forward and reverse primer at 5 µM, 2 µL adapter-ligated DNA at 0.12 ng/µL, and 1 µL H_2O. Working under the assumption that Illumina primers have been validated as unbiased by Illumina, we tested all forward (iTru5) primers with Illumina D701 as the reverse primer, and we tested all reverse primers (iTru 7) with Illumina D501 as the forward primer. We ran all primer tests in duplicate on an Applied Biosystems StepOnePlus (Life Technologies, Carlsbad, CA, USA) using the following conditions: 95ºC for 2 min, then 40 cycles of 95ºC for 15 s, and 60ºC for 1 min. Because we needed to run multiple plates of qPCR to test all of the primers, we included the iTru5 set 2 primer A (iTru5_02_A) and the iTru7 set2 primer 1 (iTru7_02_01) on all plates to provide a baseline reference for iTru5 or iTru7 primer performance. We determined the threshold cycle (C_T) using the default settings of the StepOnePlus, we averaged C_T values from replicate runs, and we calculated Delta C_T for each iTru primer using two approaches. First, we evaluated the relative performance of all iTru5 and iTru7 primers by subtracting the C_T of the iTru5 or iTru7 primer being tested from the average C_T of all iTru5 or iTru7 primers. Second, we evaluated the performance of all iTru5 and iTru7 primers by subtracting the baseline reference C_T of iTru5_02_A from the C_T of the iTru5 primer being tested and by subtracting the baseline reference C_T of iTru7_02_01 from the C_T of the iTru7 primer being tested. We

288 expected that unbiased primers would not deviate from the average and/or baseline performance
289 by more than 1.5 PCR cycles (>1.5 C_T), a value that should encompass the stochasticity seen
290 between independent PCR reactions as a result of small, unavoidable primer concentration and
291 other amplification performance differences.
292
293 DNA samples
294 To test the performance of both our Y-yoke adapters and the iTru system in a variety of library
295 preparation scenarios, we prepared libraries from DNA of various types and quality. As a
296 simple, known source of control DNA, we used *Escherichia coli* k-12 strain MG1655 (hereafter
297 *E. coli*; Roche 454, Inc.) which has a high-quality genome sequence available (GenBank
298 accession NC_000913; 4.6 Mb) and which is commonly used for quality control of sequencing
299 libraries. To examine how our iTru system performed with DNA of varying quality and
300 complexity, we also prepared iTru libraries from DNA that we isolated from a diverse array of
301 six species (three sharks, a tarantula, jellyfish, and coral). We isolated each of these DNA
302 sources using a variety of techniques commonly used in many labs, including commercial kits,
303 salting out, or CTAB Phenol-Chloroform extraction (Table 4; also see Supplemental File 1 for
304 additional details about testing iNext). We felt that these samples represented the range of
305 species, sampling conditions, and DNA isolation techniques that are commonly encountered in
306 model and non-model organism studies, and the taxa we sampled included particularly
307 challenging specimens (i.e., tarantula, coral and jellyfish) that have previously performed poorly
308 with commercial library preparation kits. Before library preparation, we fragmented *E. coli*
309 genomic DNA to 400-600 bp using a Covaris S2 (Covaris, Woburn, MA, USA), and we

fragmented genomic DNA (normalized to 23 ng/µL) to 400-600 using the Bioruptor UCD-300 sonication device (Diagenode, Denville, NJ, USA).

Library construction

Prior to library preparation, we annealed the iTru adapter sequences to form double-stranded, Y-yoke adapters by mixing equal volumes of the iTru_R1_stub and iTru_R2_stub_RCp oligos at 100 µM, supplementing the mixture with 100 mM NaCl, heating the solution to 98°C for 2 min in a thermal cycler, and allowing the thermal cycler to slowly cool the mixture to room temperature (Supplemental File 7).

We prepared iTru libraries from *E. coli* using kits, reagents, and protocols from Kapa Biosystems (Wilmington, MA, USA), with minor modifications to the manufacturer's instructions. The largest change we made was to ligate the universal iTru adapter stubs (Table 3, Supplemental File 4) to the 3'-adenylated (i.e., +A) DNA fragments, and then use the iTru5 and iTru7 primers with TruSeqHT indexes for limited-cycle amplification. For the eukaryotic libraries, we further modified the manufacturer's instructions by using half-volume reaction sizes with the following two changes. First, we used an inexpensive alternative to commercial SPRI reagents (Sera-Mag Speedbeads Thermo-Scientific, Waltham, MA, USA; see Supplemental File 8) in all cleanup reactions at a ratio of 2.86:1. Second, after adapter ligation, we performed one post-ligation cleanup followed by dual-SPRI size selection using 55 µL of PEG/NaCl and 25 µL of Speedbeads. We outline step-by-step methods for this approach in Supplemental File 9.

Sequencing

We quantified libraries using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and Kapa qPCR (Kapa Biosystems), checked for index diversity (Supplemental File 10), and then normalized and pooled all libraries at 10 nM (Supplemental File 11). We also ensured the quality of library pools by running 1 µL on a Bioanalyzer High Sensitivity chip (Agilent Technologies, Santa Clara, CA, USA). We combined the iTru and iNext *E. coli* library pools (Supplemental File 1) with samples from other experiments, and we sequenced the combined pools using a single Illumina MiSeq v2 500 cycle kit (PE250). We combined the eukaryotic libraries with additional TruSeq libraries from other experiments and sequenced these on a separate run of Illumina MiSeq v2 500 cycle kit to produce PE250 reads.

Sequence Analysis

After sequencing, we demultiplexed reads using Illumina software (bcl2fastq v 1.8 – 2.17; Illumina 2013). We then imported reads to Geneious 6.1.7 – R9.0.4, and trimmed adapters and low-quality bases (<Q20). We removed reads with inserts of <125 bases prior to all downstream analyses. We mapped *E. coli* reads back to NC_000913 using the Geneious mapper (fastest setting, single iteration). We assembled reads from the eukaryotic libraries using the Geneious assembler (fastest setting), and we extracted contigs of 250 to 450 bp from eukaryotic libraries of tarantula, jellyfish, and coral for downstream microsatellite searches using msatCommander 1.0.8 (Faircloth 2008). We also used PAL_FINDER v0.02.03 (Castoe *et al.* 2012) to enumerate microsatellites within read-pairs that had inserts ≥250 bases. Finally, we extracted contigs of approximately 17 kb from the shark libraries, and we used MEGA-BLAST searches to determine which of these contigs represented shark mtDNA genomes (Díaz-Jaimes *et al.*, 2016), similarly

354 with the coral, but an 18 kb fragment contained the mtDNA genome (Del Rio-Portilla *et al.*,
355 2016)
356
357 Larger-scale Tests
358 Following initial validation of the iTru primers and the utility of the iTru library preparation
359 approach, we put the iTru system into an extensive test phase in which we routinely used this
360 approach for library construction within our own labs while we also made all components of the
361 iTru system available to dozens of other labs. To demonstrate the utility of our approach across
362 a variety of projects, we analyzed read count data from four of these studies (n=576 libraries)
363 that used the iTru system as part of a workflow for target enrichment of ultraconserved elements
364 (UCEs; Faircloth *et al*. 2012). These included 90 iTru libraries prepared by our group from
365 cichlid fishes (McGee *et al*. 2016), 183 iTru libraries prepared by a second group (R. Harrington,
366 personal communication) from carangimorph fishes, 100 iTru libraries prepared by a third group
367 (M. Branstetter, personal communication) from ants, and 203 iTru libraries prepared by our
368 group from birds. For the bird libraries, we prepared one batch of standard Illumina libraries
369 (n=10) and 2 batches of iTru libraries (n=203), which allowed us to look at sample-to-sample
370 differences in read counts returned from standard Illumina libraries relative to our iTru libraries.
371 One of the two batches of iTru libraries (n=92) combined standard Illumina primers (D5xx;
372 which we used on *E. coli*) on the P5 side with iTru7 primers on the P7 side. The second batch
373 (n=111) combined iTru5 primers on the P5 side with iTru7 primers on the P7 side. The first
374 batch allowed us to assess iTru7 performance separate from that of iTru5, while the iTru7+iTru5
375 libraries allowed us to assess performance of the full iTru system relative to all other
376 combinations. For all remaining libraries within the other projects, each group followed the

protocols for iTru library preparation described above using combinations of only iTru5 and iTru7 primers.

Following library preparation and PCR amplification, each laboratory combined all libraries into equimolar pools containing 8-12 libraries and followed a standardized protocol for target enrichment of UCE loci (http://ultraconserved.org). After enrichment, each group used a Bioanalyzer to determine the insert size of enriched libraries and, to reduce the variance in number of reads sequenced from each pool, quantified pools using a commercially available qPCR kit (KapaBiosystems). Prior to sequencing, all research groups used the average fragment size distribution and qPCR concentration of each pool to produce an equimolar, project-specific pool-of-pooled-libraries for sequencing with a final concentration of 10 nM. We sequenced the enriched cichlid and carangimorph libraries using different, partial runs of PE150 sequencing on an Illumina NextSeq, the ant libraries using one lane of PE125 sequencing on an Illumina HiSeq 2500, and the bird libraries using two lanes of PE150 sequencing on an Illumina HiSeq 1500 (Rapid Run Mode). For the carangimorph fish libraries, we wanted each sample to receive 0.5% of the total number of reads in the NextSeq run. For all other libraries, we wanted each library to receive 1% of the total number of reads. After sequencing, we computed the average number of raw reads returned per sample, the 95% confidence interval (95 CI) of reads returned per sample, and the percentage of reads returned per sample.

Results

Quantitative PCR

Almost all iTru primers (158/160 iTru7's and 48/48 iTru5's) had average C_T values within 1.5 cycles of both the average ΔC_T and the baseline ΔC_T (Figure S8; Supplemental File 12),

suggesting that our iTru indexed amplification primers amplify successfully (98.7% success for iTru7; 100% success for iTru5) and perform similarly to one another. There were two iTru7 primers that failed to amplify during their initial tests, iTru7_401_07 and iTru7_209_04. We rehydrated a new plate of primers and retested iTru7_401_07, which amplified normally (C_T = 19.4, ΔC_T (average) = -0.7; ΔC_T (baseline) = 1.1) during the retest.

E. coli iTru Libraries

The iTru libraries we prepared from *E. coli* returned similar numbers of reads from each iTru library, averaging 973,008 reads per sample (95 CI: 161,044; Fig S9, Supplemental File 13). Each library contained >400,000 high quality reads that covered >99.99% of the known *E. coli* genome sequence. These results suggest that our iTru library preparation process produces valid constructs for Illumina sequencing, and that iTru dual-indexed libraries pooled at equimolar ratios return roughly similar amounts of sequence data (Fig. S9), although we combined libraries at equimolar ratios prior to sequencing using fluorometry, which can result in some variation around the targeted read number for each library.

Eukaryotic Species iTru Libraries

We successfully sequenced all eukaryotic libraries prepared using the iTru system and the libraries returned an average of 1,806,440 reads per sample (95 CI: 743,337; Table 4). Using a genome skimming approach, we sequenced the mitogenomes of the shark and coral samples to an average coverage of 33x and 50x, respectively. We used the contig assemblies from our tarantula, jellyfish, and coral samples to design primers pairs targeting >100 microsatellite loci in each taxon. Although the variance in the number of sequencing reads returned per library was

higher among these samples than the *E. coli* libraries, these results demonstrate that the iTru system can be used to prepare libraries from DNA of different organisms extracted using different purification approaches, even DNA that produced very poor results with commercial kits (data not shown).

Larger-scale Tests

Our beta test allowed us to collect sequence data from many different iTru5 and iTru7 primers used to index a variety of iTru libraries from fishes, ants, and birds. Few of the libraries that we or others prepared using the iTru system showed large differences in the desired number of reads sequenced when compared to libraries having Illumina-only adapters/index sequences when viewed in aggregate (Fig. S10) or on an index-by-index basis across projects (Figs. S11-14; Supplemental File 14). The iTru primer combinations that sometimes returned a lower number of reads for a particular library in a particular project did not show this behavior in other studies (e.g., compare iTru7_402_07 in Fig. S13 versus S14), suggesting that the reduction in read numbers results from particular library preparation, pooling, enrichment, and quantification practices for specific samples (i.e., specific experimental errors, library preparation methods, or sample-index interactions) rather than inherently bad iTru indexes/primers.

Discussion

Our results show that the iTru universal adapter stubs and iTru primers can be used to produce genomic libraries for a variety of purposes. The low variance in C_T values among iTru5 and iTru7 primers demonstrates that the different index sequences have minimal effect on the libraries, and our results from real-world tests demonstrate that the iTru system works well with

446 DNA from different extraction methods and of differing quality, quantity, and copy number.
447 The results we present from DNA libraries prepared using the iTru system in our and others
448 laboratories show that the approach easily scales to hundreds of libraries prepared, pooled, and
449 sequenced in a single lane, ultimately producing information consistent with the variety of
450 Illumina library techniques we have employed to obtain similar data (Crawford *et al.* 2012,
451 McCormack *et al.* 2013, Smith *et al.* 2014).

452 After testing the iTru system in several labs, we made several changes in our approach.
453 The most significant of these were: 1) to modify our original naming scheme so that researchers
454 can easily identify sets of iTru7 primers that are compatible or incompatible with TruSeq
455 indexes, and 2) to increase the amount of iTru5 and iTru7 aliquoted into plates after oligo
456 synthesis (from 0.625 nmol to 1.25 nmol), which reduced library amplification failures that
457 resulted from improper hydration of low-quantity primers in specific wells of plates. The
458 naming scheme and concentrations used in all supplemental files and the naming scheme we
459 used in the Methods section reflect these changes to minimize confusion. After making these
460 changes, we and others have successfully produced libraries and sequencing reads from the vast
461 majority of iTru5 and iTru7 primers detailed in the supplemental files, and we have no evidence
462 suggesting that any of the primer sequences will not work correctly. The original sets of iTru7
463 primers (sets 00 – 13) exist, but they have mixed compatibility with Illumina indexes, thus we
464 encourage beta users to exhaust those primers quickly and adopt the new sets.

465 It is important to note that the iTru5 and iTru7 primers are grouped into "balanced" sets
466 of 8 or 12 to minimize problems of index base diversity during sequencing. Index balance
467 problems arise because of the way Illumina platforms detect bases during the sequencing run
468 (Illumina 2016b), and the main issues associated with unbalanced base composition are

experienced when relatively few samples are sequenced or when a small number of libraries with unbalanced sequence tags take up a large fraction of the sequencing run. We modeled the original four color-scheme used in HiSeq and MiSeq instruments, because the color scheme used in NextSeq and MiniSeq instruments had not yet been released. Using an entire group of eight iTru5 and 12 iTru7 indexed primers within a sequencing pool where each library is present in equal proportion ensures balanced base representation during the index sequence read(s). Generally, when researchers multiplex more than one group of eight iTru5 or 12 iTru7 indexed primers, base diversity is even more balanced, although it is always a good idea to check the balance of sequencing tags in all sequencing runs (i.e., use Supplemental File 10). When less than a whole set of primers (i.e., <8 iTru5 primers or <12 iTru7 primers) are used, or if very few libraries will dominate the percentage of reads within a run, it becomes critical to ensure the tags are sufficiently diverse (i.e., use Supplemental File 10; which now includes separate calculations of base diversity for both color schemes).

All of the iTru oligonucleotides make use of a single phosphorothioate bond between the penultimate and 3' base. Phosphorothioate linkages protect the 3' end of oligonucleotides from some forms of nuclease activity (Ekstein 1985, Skerra 1992) such as those introduced by some DNA ligases and polymerases (exonuclease activity is a common contaminant of ligases and an intrinsic activity of proofreading polymerases), but phosphorothioate linkages add a modest cost to each primer (~$3 USD per phosphorothioate linkage). Phosphorothioate linkages are also chiral, so only 50% of synthetic molecules receive protection per linkage, while the other 50% remain susceptible to nuclease activity (Eckstein, 1985). Adding a second phosphorthioate bond can reduce the proportion of unprotected molecules by 50% (thus 75% would be protected and 25% would remain susceptible). Illumina and other vendors often include three or more

492 phosphorothioate linkages at the 3' end of their oligonucleotides to ensure that a large fraction
493 are protected from nuclease activity. We include only a single phosphorothioate linkage in our
494 iTru oligo designs because if we lose the 3' base, we would rather lose the rest of the molecule
495 instead of rescuing the remaining part of it, which may not function appropriately. This strategy
496 also reduces costs associated with synthesizing the oligonucleotides, although others may prefer
497 to incorporate additional phosphorothioate linkages (e.g., two phosphorotioate linkages would
498 lead to 50% fully protected oligonucleotides and 25% that only lose a single 3' base).
499
500 Who should adopt this method?
501 Researchers who need higher plexity of their Illumina library preparations or who have not yet
502 invested heavily in any other method will likely find our approach attractive. It has a low cost of
503 entry and significant flexibility (see below). The more types of libraries, projects, and samples
504 researchers use, the quicker they will recoup the cost of switching and see savings. Additionally,
505 researchers using MK-2010 to construct libraries with inserts >200 bp, particularly those inserts
506 ≥500 bp, are likely to benefit from using a Y-yoke adapter.
507 Researchers already invested in and using other methods with good success, such as the
508 MK-2010 or F-2011 approaches, may wonder if it is worthwhile to switch. We suggest that it
509 would be reasonable to continue using the MK-2010 and/or F-2011 methods if these are already
510 being used successfully; for these labs, we simply provide some alternative adapters and primers
511 that could be used once existing stocks of MK-2010 and/or F-2011 adapters and primers are
512 exhausted or when new projects requiring unique or larger numbers of uniquely tagged samples
513 are encountered.
514

515 iNext

516 In addition to the iTru adapters and primers we designed and tested, we have developed a
517 universal adapter stub and sets of primers (iNext; Supplementary File 1) that are compatible with
518 the Illumina Nextera system and the original 8x12 Nextera indexes, though they are not
519 compatible with all of the subsequent Nextera indexes. As noted in the methods, both iNext and
520 iTru make use of slightly different subsets of the tags identified by Faircloth and Glenn (2012),
521 and the indexed primer sets and numbering approaches are independent between iNext and iTru
522 (e.g., iNext5_01_A does not have the same sequence tag as iTru5_01_A). Thus, researchers
523 should use the tag sequence or tag number from Faircloth and Glenn (2012) or the tag sequences
524 themselves to determine which indexes are equivalent (e.g., iNext7_07_06 uses tag 113
525 [AGCTAAGC] as does iTru7_203_10; these should not be combined into a single sequencing
526 pool). Although we demonstrate it is possible to combine iNext and iTru libraries within the
527 same MiSeq run (Supplemental File 13; the iNext and iTru *E. coli* data come from a single
528 MiSeq run), and while we have subsequently added iNext or Nextera libraries in limited
529 quantities to several of our iTru library pools run on the MiSeq, we are skeptical that other
530 researchers should or will do this routinely. If researchers want to combine iNext and iTru
531 libraries on a regular basis, it would be worthwhile to run additional experiments and to screen
532 and sort the tags to compile sets with numbering that is consistent, thus facilitating pooling
533 between the two systems.

534

535 Troubleshooting (Finding What Tag Combinations Were Actually Used)

536 Although all researchers endeavor to conduct mistake-free experiments, foul-ups are certain to
537 occur. In addition to simple record-keeping errors, a very common mistake is flipping the

orientation of one of the strip tubes containing iTru primer aliquots. Thus, it is critical to have the capacity to quickly and easily determine what index sequences and combinations are present within a sequencing run. We have developed a small and fast python program (Supplementary File 17) that can count the indexes within a file of reads that were not assigned to specific samples during demultiplexing (i.e., the undetermined reads from bcl2fastq).

Other Applications and Future Modifications

It is possible to use the iTru system for a variety purposes beyond what we describe here. For example, we have used the iTru system for making RNAseq libraries using Kapa library kits, but any approach that yields double-stranded template molecules with a single adenosine can be used with no significant modifications to what we have described. One of the attractive features of our system is that it separates the primers and stubs into more manageable units. We have also used several of the approaches described above to modify the iTru system for use with amplicon sequencing and RADseq studies. In subsequent *Adapterama* papers, we use these same iTru primers with different adapter stubs to construct double- to quadruple-indexed amplicon libraries (Glenn *et al*. 2016a), double-digest restriction-site associated DNA (RAD; Glenn *et al*. 2016b), and RAD-capture (Hoffberg *et al*. 2016) libraries. All of these extensions facilitate library preparation, sequencing, and bioinformatic processing of these types of data while also significantly reducing costs.

Having separate primers and adapter stubs simplifies and reduces costs associated with modification or swapping out of the universal Y-yoke adapters (Table 3, Supplemental Files 4, 15), creating opportunities for further research and protocol development. For example, if researchers wanted to optimize library preparation for low levels of input DNA, then

implementing an adapter stub in a stem-loop configuration [cf. New England Biolabs (NEB) Next Ultra] would be worth investigating. Similarly, adapters containing uracils that are broken at the uracil sites by USER (NEB M5505) or uracil-DNA-glycosylase (UDG; e.g., NEB M0280) plus APE 1 (e.g., NEB M0282) facilitate a variety of designs with potentially beneficial characteristics worth exploring, especially for mate-pair libraries. However, given recent advances in commercial kits that reduce buffer exchanges and increase efficiency (e.g., Kapa BioScience's Hyper and HyperPlus preps which require as little as 1 ng of input DNA), it is likely that the use of such high efficiency approaches combined with the iTru adapters and primers will be sufficient for the vast majority of applications where samples derive from ≥1000 eukaryotic cells.

Summary

We describe an approach that uses a single universal adapter stub and relatively few PCR primers to produce many Illumina libraries. The approach allows multiple researchers to have unique primer sets so that libraries from individual researchers can be pooled without worrying about tag overlap. These primers can also be used with a variety of other application-specific adapters described in subsequent *Adapterama* papers for amplicon and RADseq libraries (Glenn *et al*. 2016 a, b; Hoffberg *et al*. 2016). By modularizing library construction, researchers are free to focus on the development of new application-specific tags. Taking advantage of the many available tags also creates opportunities for low-cost experimental optimization attempts. Although the adapters and primers we describe are specific to Illumina, many of the ideas can easily be extended to Ion Torrent, Pacific Biosystems, Oxford Nanopore, and other sequencing platforms (c.f. Glenn *et al*. 2007).

584

585 **Acknowledgements**

586 Oligonucleotide sequences © 2007-2016 Illumina, Inc. All rights reserved. Derivative works
587 created by Illumina customers are authorized for use with Illumina instruments and products
588 only. All other uses are strictly prohibited. We thank Rigoberto Delgado Vega, Yann Henaut,
589 Salima Machkour M'Rabet, Fausto Valenzuela, Eduardo Balart, David Paz, Carolina Galvan,
590 Liza Gomez Daglio, and Pindaro Díaz Jaimes for providing samples, Erin Lipp for generously
591 sharing laboratory space and equipment, Rahat Desai for helpful edits, and our colleagues at the
592 Georgia Genomics Facility for conducting laboratory analyses. We thank Richard Harrington,
593 Matt Friedman, and Thomas Near (carangimorph fishes), and Michael Branstetter, John
594 Longino, and Phil Ward (ants) for allowing us to use read-count information from their
595 respective studies. This work was partially supported by DEB-1242241, DEB-1242260, DEB-
596 1136626, DEB-1146440, DGE-0903734, and OISE 0730218 from the U.S. National Science
597 Foundation, SAGARPA-FIRCO (grant RGA-BCS-12-000003), UCMEXUS (grant CN-13-617),
598 and CB-CONACYT (grant 157993). Finally, we acknowledge and thank Lisa Ortuno (deceased)
599 for her enthusiastic support of this work; our world was enriched while she shared it with us.

600

601 **Conflicts of Interest**

602 The authors declare competing interests. The EHS DNA lab and Georgia Genomics Facility
603 provide oligonucleotide aliquots and library preparation services at cost, including some
604 oligonucleotides and services that make use of the adapters and primers presented in this
605 manuscript. The information we present allows all researchers to synthesize the
606 oligonucleotides at any vendor of their choice, follow or modify the library preparation

607 techniques we have included, and freely publish results simply with proper attribution of this
608 paper and Illumina®™.

609

610

611 **Data Availability**

612 Supplemental figures, files, and protocols are available from Dryad (doi: xxxx). Sequence reads
613 generated as part of this study for the main text and in Supplementary File 1 are available from
614 the NCBI (PRJNA321721). A video presentation giving an overview of the iTru system that is
615 available from: https://www.youtube.com/watch?v=sRATXu2Ff08.

616

617 **Author Contributions**

618 TCG and BCF conceived iTru and iNext; TCG and BCF designed iNext and the tags used; RN
619 designed the iTru stubs and initial primers; RN, TJK, JWF Jr., TWP, JLA, and JKM conducted
620 labwork; TCG, RN, SL, and BCF analyzed data; TCG and BCF drafted most text; TJK, JWF Jr.,
621 TWP, KEB, and SLH made figures, tables, and supporting materials; FJGL, MARP, KR, JLA,
622 JKM, SA, RR, MA, and MB provided samples, funding and additional data analysis; all authors
623 edited and commented on the manuscript.

References

Adey A, Morrison HG *et al.* (2010) Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. *Genome Biology*, **11**, R119.

Aird D, Ross MG, Chen WS *et al.* (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, **12**, R18.

Aljanabi SM, Martínez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research,* 25:4692-4693.

Ansorge WJ (2009) Next-generation DNA sequencing techniques. *Nature Biotechnology,* **25**, 195-203.

Bentley DR, Balasubramanian S, Swerdlow HP *et al.* (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, **456**, 53-59.

Binladen J, Gilbert MTP, Bollback JP *et al.* (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE*, **2**, e197.

Castoe TA, Poole AW, de Koning APJ *et al.* (2012) Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLoS ONE*, **7**, e30953.

Craig DW, Pearson JV, Szelinger S *et al.* (2008) Identification of genetic variants using bar-coded multiplex sequencing. *Nature Methods*, **5**, 887-893.

Crawford NG, Faircloth BC, McCormack JE *et al.* (2012) More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biology Letters,* **8**, 783-6.

Díaz-Jaimes P, Hinojosa-Alvarez S, Sánchez-Hernández X, Hoyos-Padilla M, García-De-León FJ (2016). The complete mitochondrial DNA of white shark (*Carcharodon carcharias*) from Isla Guadalupe, Mexico. *Mitochondrial DNA*, **27**, 1281-1282.

Del Rio Portilla MA, Vargas Peralta CE, Paz García DA, Lafarga De La Cruz F, E. Balart Páez E, García De León FJ (2016) The complete mitochondrial DNA of endemic Eastern Pacific coral (*Porites panamensis*). *Mitochondrial DNA*, **27**, 738-739.

Eckstein F (1985). Nucleoside phosphorothioates. *Ann. Rev. Biochem.*, **54**, 367-402.

Faircloth BC (2008). MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources,* **8**, 92-94.

Faircloth BC, Glenn TC (2012) Not all sequence tags are created equal: Designing and validating sequence identification tags robust to indels. *PLoS ONE*, **7**, e42543

Faircloth BC, McCormack JE, Crawford NG, *et al.* 2012. Ultraconserved elements anchor thousands of genetic markers for target enrichment spanning multiple evolutionary timescales. *Systematic Biology*, **61**, 717-26.

Fisher S, Barry A, Abreu J, Minie B, Nolan J, et al. 2011. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biology* 12:R1.

Galván-Tirado C, Hinojosa-Alvarez S, Diaz-Jaimes P, Marcet-Houben M, García-De-León FJ (2016). The complete mitochondrial DNA of the silky shark (*Carcharhinus falciformis*). *Mitochondrial DNA*, **27**, 157-158.

Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, **11**, 759-769.

667 Glenn TC (2016) 2016 Update to the field guide to next-generation DNA sequencers.
668 http://www.molecularecologist.com/next-gen-fieldguide-2016/ accessed 04/15/2016.

669 Glenn TC, Jones KL, Lance SL *et al.* 2007. Source tagging and normalization of DNA for
670 parallel DNA sequencing. U.S. Provisional Patent 60/909,010 filed March 30, 2007.

671 Glenn TC, Pierson TW, Kieran TJ, *et al.* (2016a) Adapterama II: Universal amplicon
672 sequencing in Illumina platforms (TaggiMatrix). Submitting to Molecular Ecology
673 Resources and bioRxiv

674 Glenn TC, Pierson TW, Kieran TJ, *et al.* (2016b) Adapterama III: Quadruple-indexed triple-
675 enzyme RADseq libraries (3RAD). Submitting to Molecular Ecology Resources and
676 bioRxiv

677 Greigite (2009). adapter illustration.pdf available at
678 http://seqanswers.com/forums/showpost.php?p=7629&postcount=36 accessed 27 July,
679 2014.

680 Head SR, Komori HK, LaMere SA *et al.* (2014) Library construction for next-generation
681 sequencing: Overviews and challenges. *BioTechniques*, **56**, 61-77.

682 Hoffberg S, Kieran TJ, Catchen JM, *et al.* (2016) Adapterama IV: Sequence capture of dual-
683 digest RADseq libraries with identifiable duplicates (RADcap). Submitted to Molecular
684 Ecology Resources and bioRxiv: http://biorxiv.org/content/early/2016/04/20/044651.

685 Hoffmann C, Minkah N, Leipzig J *et al.* (2007) DNA bar coding and pyrosequencing to identify
686 rare HIV drug resistance mutations. *Nucleic Acids Research*, **13**, e91.

687 Illumina, 2008. Multiplexed sequencing with the Illumina Genome Analyzer system. Pub. No.
688 770-2008-011. Available at:

http://www.illumina.com/documents/products/datasheets/datasheet_sequencing_multiplex.pdf, accessed 28 April 2016.

Illumina, 2012. Frequently asked questions; Sequencing; Can I run TruSeq HT libraries with Nextera libraries on the same lane or flow cell? Available at: http://support.illumina.com/sequencing/sequencing_kits/truseq_dna_ht_sample_prep_kit/questions.ilmn, accessed 28 April 2016.

Illumina, 2013. bcl2fastq conversion user guide. Illumina Proprietary Part #15038058 Rev B, March 2013. http://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/bcl2fastq/bcl2fastq_letterbooklet_15038058brpmi.pdf, accessed 7 June 2016.

Illumina, 2016a. Indexed sequencing overview guide. Illumina Proprietary Document #15057455v02, March 2016. http://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/indexed-sequencing-overview-guide-15057455-02.pdf, accessed 2 June 2016.

Illumina, 2016b. Nextera low plex pooling guidelines. Technical note: DNA Analysis. Pub. No. 770-2011-044, current as of 24 March 2016. Available at: http://www.illumina.com/documents/products/technotes/technote_nextera_low_plex_pooling_guidelines.pdf, accessed 28 April 2016.

Kircher M, Sawyer S, Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, **40**, e3.

709 Kozerewa I, Ning Z, Quail MA *et al.* (2009) Amplification-free Illumina sequencing-library
710 preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nature*
711 *Methods*, **6**, 291-295.

712 McCormack JE, Harvey MG, Faircloth BC *et al.* (2013) A phylogeny of birds based on over
713 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS ONE*, **8**,
714 e54848.

715 McGee MD, Faircloth BC, Borstein SR, Zheng J, Hulsey CD, Wainwright PC, Alfaro ME
716 (2016). Replicated divergence in cichlid radiations mirrors a major vertebrate innovation.
717 *Proceedings of the Royal Society of London: Biological Sciences*, **283**, 20151413.

718 Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed
719 target capture and sequencing. Cold Spring Harbor Protocol, 2010, pdb prot5448

720 Meyer M, Stenzel U, Myles S *et al.* (2007) Targeted high- throughput sequencing of tagged
721 nucleic acid samples. *Nucleic Acids Research*, **35**, e97.

722 Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for
723 multiplexed target capture. *Genome Research*, **22**, 939-946.

724 Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW (1996) Quantitative
725 phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding
726 strategy. *Nature Genetics*, **14**,450-456.

727 Skerra A (1992) Phosphorothioate primers improve the amplification of DNA sequences by
728 DNA polymerases with proofreading activity. *Nucleic Acids Research*, **20**, 3551-3554.

729 Smith BT, Harvey MG, Faircloth BC, *et al.* (2014) Target capture and massively parallel
730 sequencing of ultraconserved elements (UCEs) for comparative studies at shallow
731 evolutionary time scales. *Systematic Biology*, **63**, 83-95.

732 Syed F, Grunenwald H, Caruccio N (2009). Next-generation sequencing library preparation:
733 Simultaneous fragmentation and tagging using *in vitro* transposition. *Nature Methods*
734 *Application Note 6*. http://www.nature. com/nmeth/journal/v6/n11/full/nmeth.f.272.html.
735 Tautz D, Ellegren H, Weigel D (2010) Next generation molecular ecology. *Molecular Ecology,*
736 **19 s1**, 1-3.
737

Box 1 – Glossary

adapters – oligonucleotides of known sequence that are ligated onto the ends of nucleic acids for the purpose of further manipulation or NGS library construction. In this paper we will make use of double-stranded DNA adapter **stubs** (see below).

barcodes – see index or tag; this term is also used to mean a DNA sequence that can be used to identify the taxon from which a sample derives, thus we avoid using this ambiguous term.

cluster – a group of molecules on an Illumina flow cell that have been clonally amplified via bridge PCR or newer approaches (i.e., all molecules in a cluster are replicates of a single starting molecule from an Illumina library).

demultiplex – to separate pooled (multiplexed) sample information into their constituent parts (i.e., assign reads to specific samples)

identifying sequences – see index or tag.

index or tag – a short, unique sequence of DNA added to samples so they can be pooled and sequenced in parallel, with each resulting sequence containing information to identify the source sample. Some authors and companies refer to such sequences as barcodes or molecular identifiers (MIDs). We use "Illumina index" when referring to specific sequences designed by Illumina, "tag" when specifically referring to sequences from Faircloth and Glenn (2012), and "index" when generically referring to identifying sequences in adapters and primers compatible with Illumina instruments.

Index Read 1 – the DNA sequence obtained from the 2nd Illumina sequencing reaction, yielding the i7 index sequence, which is placed into the header of Read 1 and Read 2 (if present)

Index Read 2 – the DNA sequence obtained from the 3rd Illumina sequencing reaction, yielding the i5 index sequence, which is placed into the header of Read 1 and Read 2 (if present)

i5 index – the second indexing position introduced by Illumina, obtained by index Read 2, which is the 3rd read of a cluster made by Illumina instruments.

i7 index – the original indexing position used in Illumina sequencing, obtained by index Read 1, which is the 2nd read of a cluster made by Illumina instruments.

iNext – dual-index library preparation methods presented herein that are compatible with Illumina Nextera libraries.

iTru – dual-index library preparation methods presented herein that are compatible with Illumina TruSeq libraries.

library – a population of molecules with adapters on each end of each molecule to facilitate sequencing.

MID – molecular identifier, term commonly used with 454 sequencing; see index or tag.

multiplex – samples that are pooled together and processed or sequenced all at once.

P5 – an engineered DNA sequence that is: 1) incorporated into adapters of Illumina libraries for bulk amplification of library molecules and 2) manufactured as oligonucleotides grafted onto the surface of Illumina flow cells and used for clonal amplification of library molecules, and priming the 3rd sequencing reaction on MiSeq and HiSeq ≤2500 instruments.

P7 – an engineered DNA sequence that is: 1) incorporated into adapters of Illumina libraries for bulk amplification of library molecules and 2) manufactured as oligonucleotides grafted onto the surface of Illumina flow cells and used for clonal amplification of library molecules.

paired-end reads – DNA sequences obtained from sequencing each strand of DNA templates within clusters (see Fig. 4).

primers – single-stranded oligonucleotides used to initiate strand elongation for sequencing or amplification

Read 1 – the DNA sequence obtained from the 1st Illumina sequencing reaction, obtained as a fastq file with headers that contain data from indexing reads 1 and 2.

Read 2 – the DNA sequence obtained from the 4th Illumina sequencing reaction, obtained as a fastq file with headers that contain data from indexing reads 1 and 2.

sequence diversity – the base composition of nucleotides across all clusters being sequenced at any given base position. Illumina sequencing requires sequence diversity for successful determination of a base call.

stubs – short universal adapters that are formed by annealing two oligonucleotides together (Illumina Read1 and Read2 sequences) and attaching that double-stranded product to template DNA via ligation. In the iTru strategy, y-yoke adapter stubs are comprised of oligonucleotides with the Read1 and Read2 sequences.

y-yoke – an adapter that is formed from two oligonucleotides that are complementary on only one end to form a product that is double-stranded at one end, but single-stranded at the other end.

800 **Figure Legends**

801

802 **Figure 1. Overview of iTru library construction.** Sheared DNA from the organism of interest
803 (black) is used as input for iTru library preparation process. The input DNA is end-repaired and
804 a single adenosine (A) overhang (not shown) is added to the 3' end (see Figs. 2, 3 for details). Y-
805 yoke adapter stubs, which have annealed complementary regions (orange) of the Read 1 (R1,
806 purple) and Read 2 (R2, red) adapters, a 3' thymidine (T) overhang (not shown), and are
807 phosphorylated (indicated with a "P" at the 5' position), are ligated to the genomic DNA.
808 During limited-cycle PCR, iTru5 and iTru7 primers anneal to the ends of the Y-yoke adapters
809 and are extended to produce full-length, double-indexed molecules (see Fig. S3 for details of
810 PCR), making them fully functional for sequencing on Illumina instruments and also adding dual
811 indexes. The P5 (maroon) and P7 (yellow-green) regions on the molecule are complementary to
812 oligonucleotides present on Illumina flow-cells, allowing for hybridization and clonal
813 amplification. The i5 (green) and i7 (light blue) indexes can be used for multiplexing. The R1
814 and R2 primer-binding sites are complementary to the sequencing primers, enabling sequencing
815 of the library molecules on the flow cell. The R1 and R2 primer-binding sites also contain
816 regions with identical sequence (shown in orange) that are used to facilitate the y-yoke adapters.
817 Thus, the full R1 and R2 sequences include the regions in orange (see Fig. 2).

818

819 **Figure 2. iTru and iNext library preparation workflows.** Here we illustrate the major steps
820 used for library construction. The process is identical for iTru and iNext, except: 1) which
821 nucleoside (A vs. C) is added to blunt, 5' phosphorylated (end-repaired) molecules, 2) which

adapter is ligated to the DNA, and 3) which primers are used for limited-cycle PCR. All steps are functionally equivalent.

Figure 3. Detailed steps for iTru library construction with relevant sequences. Starting material is sheared, double-stranded DNA (represented as X) with ragged ends. The DNA is made blunt and 5' phosphates are added (phosphates not shown). Third, a single adenosine (A) is added to each 3' end to allow for complementary hybridization of adapters. Next, stubby Y-yoke adapters with complementary ends are ligated to each end of the DNA molecule. These adapters contain both complementary and non-complementary sequences (non-complementary indicated by the gap between the top and bottom strand). These non-complementary sequences include primer-binding sites, as indicated by the colors, used in the next step. In the final step of library preparation, limited-cycle PCR is performed using two distinct primers complementary to the ends of the Y-yoke adapter (shown as iTru5 and iTru7). The primers contain unique indexes (i5 and i7, respectively, shown in color) as well as the P5 and P7 sequences (for color scheme and explanation of functions, see Figure 1). The index strand in color indicates the sequence of the primer (which is the same as the index read for i5, but the reverse complement is obtained for the i7 index read; see Figure 4). Note that iNext libraries are similar, except that cytosines are added to the template DNA (instead of adenosines), the Y-yoke adapter has single guanosine overhangs, and the Read1 and Read2 portions have different sequences (cf. Fig. S4).

Figure 4. Sequencing reads that can be obtained from the full-length, dual-indexed iTru library molecules. The top double-stranded molecule shows an iTru-library-prepared molecule. The color scheme follows Figure 1, except that the sequences derived from the complementary

ends of the adapter molecules (i.e., the portion of the y-yoke adapter that was annealed together and previously shown in orange) are illustrated in light violet and light red on the template to more clearly indicate their contiguity and are not shown on the primers (Fig S6 shows these regions in orange). The horizontal arrows indicate sequencing primers (binding to the complementary strand of the library molecules). The tip of the arrowhead indicates the 3' end of the primer and the direction of elongation for sequencing. Four sequencing reads are shown for each library-prepared molecule, with one read for each index and each strand of the genomic DNA. Reads are arranged 1 to 4 (numbered in magenta) from top to bottom, respectively. Numbering follows the order in which the reads are obtained on Illumina instruments. The arrow immediately 3' of the primers indicates the data obtained from that primer. Three of the reads use the same primers for all Illumina instruments; however, the 3rd read (Indexing Read 2) uses a different primer on the NextSeq (or MiniSeq or HiSeq models ≥3000) vs. HiSeq (models ≤2500) or MiSeq. The 3rd read on the HiSeq (≤2500) and MiSeq use the P5 primers on the surface of the flow cells, which are not full length. Thus, a short "Dark Read", which uses up reagents without collecting data, is needed to extend the primer to the i5 index (see text for more details). The same read schematic is available for iNext (Fig. S4), and Figure S7 illustrates the reads generated from libraries lacking a P5 index but sequenced using double-indexing run settings on an Illumina platform.

Table 1. Comparison of oligonucleotide numbers and costs when using varying numbers of independent tags. Cost estimates assume 2-stage library preparations and list prices from Integrated DNA Technologies, 25 nmole synthesis scale, with oligonucleotides delivered in plates. An index length of 8 nucleotides is used with an edit distance ≥3 for iTru and an edit distance ≥2 for Illumina.

Uniquely indexed Libraries	Library Type	Index Positions	Stub adapter Oligos	Long adapter oligos	Indexed Primers	Adapter cost + primer cost (US $)
96	TruSeq*	1	0	1+96	0 [2[#]]	$4,019 + $18
96	TruSeq Nano HT	2	0	8 + 12	0[§]	$4,560[§] + $0
96	iTru[a]	1	2	0	1 + 96	$45 + $1,617
96	iTru[b]	2	2	0	8 + 12	$45 + $344
384	TruSeq*	1	0	1+384	0 [2[#]]	$16,029 + $18
384	iTru[a]	1	2	0	1 + 384	$45 + $6,416
384	iTru[b]	2	2	0	16+24	$45 + $689
9216	TruSeq*	1	0	1 + 9216 [e]	0 [2[#]]	$392,049 + $18
9216	iTru[a]	1	2	0	1 + 9216[c]	$45 + $153,539
9216	iTru[b]	2	2	0	96+96	$45 + $3,333
74,304	iTru[b]	2	2	0	192 + 387	$45 + $9,915

* Original TruSeq approach with custom adapters (cf. Faircloth & Glenn 2012); kits are no longer available, but the method can be home-brewed (cf. Fisher *et al.* 2010), or the adapters can be used with reagents from TruSeq Nano kits.

[#] P5 and P7 primers are used.

[§] Price includes all library preparation reagents, not just adapters; P5 and P7 primers are included in kit.

875 [a] Libraries contain both i5 and i7 tags, but only one iTru5 primer is used for all samples, thus
876 only the i7 tags are informative and are sequenced (cost efficient with old versions of HiSeq
877 ≤2500 kits). This method is no longer recommended, but illustrates cost differences.

878 [b] Both the i5 and i7 indexes are informative and are sequenced.

879 [c] Tags of 11 nucleotides are required for 9216 tags of edit distance ≥3.

Table 2. Comparison of Nextera, iNext, iTru, and TruSeq Nano HT library preparation methods.

Library Type	Nextera	iNext	iTru	TruSeq Nano HT
Input DNA (ng)	Intact (≥50)	Sheared (≥100[#])	Sheared (≥100[#])	Sheared (≥100)
Repair ends	N/A	Yes	Yes	Yes
Add DNA overhang	N/A	C	A	A
Ligate adapter	Tagmentation	iNext stub	iTru stub	TruSeq
Limited cycle PCR primers	Nextera or iNext*	Nextera or iNext	iTru	P5 and P7
Advantages	Least time	Lower cost, high diversity	Lower cost, high diversity	Industry standard
Disadvantages	Higher cost, lower diversity, less randomness[§]	More prep. time than Nextera	More prep. time than Nextera	Higher cost, more input DNA, more prep. time; not for sequence capture

* Note, iNext primers are not specified as biotinylated, and thus will not work interchangeably with Nextera libraries that use streptavidin beads to capture/normalize/purify libraries unless biotins are added. Using unmodified iNext primers requires other purification and normalization procedures.

[§] Tagmentation does not insert adapters into the genome as randomly as shearing the DNA.

[#] Hyper Prep Plus Kits (KapaBioSciences) allow input as low as 1 ng of intact DNA.

Table 3. iTru and iNext adapter stub oligonucleotides and tagged primer sequences. All sequences are given in 5' to 3' orientation. To make it clear which portions are constant among all tagged primers, as well as to identify function, the tagged primers are given in three pieces (the invariant 5' end, the tag sequence which varies among primers, and the invariant 3'end), but the primers are obtained as a single contiguous fusion of these three pieces. Complete balanced sets of primers are available as Supplemental Files (4, 15). Adapter stub oligonucleotides must be hydrated and annealed prior to use (Supplemental File 7).

iTru					
Adapter	Stub name	Stub sequence			
	iTru_R2_stub_RCp	/5Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCAC			
	iTru_R1_stub	ACACTCTTTCCCTACACGACGCTCTTCCGATCT			
	Primer Name	**5' end**	**Tag Sequence**	**3' end**	**Tag number**
i5	iTru5_01_A	AATGATACGGCGACCACCGAGATCTACAC	ACCGACAA	ACACTCTTTCCCTA*C	tag063
	iTru5_01_B	AATGATACGGCGACCACCGAGATCTACAC	AGTGGCAA	ACACTCTTTCCCTA*C	tag134
i7	iTru7_01_01	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC	AGTGACCT	ATCTCGTATGCCGTCTTCTGCTT*G	tag132
	iTru7_01_02	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC	AACAGTCC	ATCTCGTATGCCGTCTTCTGCTT*G	tag008

iNext					
Adapter	Stub name	Stub sequence			
	iNext_R2_stub_RCp	/5phos/TGTCTCTTATACACATCTCCGAGCCCACGAGAC			
	iNext_R1_stub	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG			
	Primer Name	**5' end**	**Tag Sequence**	**3' end**	**Tag number**
i5	iNext5_01_A	AATGATACGGCGACCACCGAGATCTACAC	GACACAGT	TCGTCGGCAGCGTC	tag317
	iNext5_01_B	AATGATACGGCGACCACCGAGATCTACAC	GCATAACG	TCGTCGGCAGCGTC	tag348
i7	iNext7_01_01	CAAGCAGAAGACGGCATACGAGAT	TCACCTAG	GTCTCGTGGGCTCGG	tag458
	iNext7_01_02	CAAGCAGAAGACGGCATACGAGAT	CAAGTCGT	GTCTCGTGGGCTCGG	tag172

Table 4. Results from initial iTru library preparation and sequencing tests of DNA from sharks and challenging non-model organisms. The Illumina i7 index sequences where used in these tests.

sample ID	Common name	Genus	species	DNA extraction method	i7 Index ID	raw index count	number of read pairs	Primary objective	reads for assembly or microsat scan[1]	putative mtDNA contig size in bp (mean coverage)	Microsat identified
MaF 5	white shark	*Carcharodon*	*carcharias*	Protocol 1	705	1,930,539	1,805,638	mtDNA	1,722,562	17,103 (46x)[3]	-
MaF 19	white shark	*Carcharodon*	*carcharias*	Protocol 2	707	2,075,236	1,927,792	mtDNA	2,003,858	17,138 (31x)[3]	-
MaF 10	silky shark	*Carcharhinus*	*falciformis*	Protocol 1	706	1,438,468	1,358,550	mtDNA	1,800,534	17,285 (22x)[4]	-
MaF 1	Tarantula cannonball	*Brachypelma*	*vagans*	Protocol 1	701	985,171	934,406	msats	80,790	-	563
MaF 16	jellyfish	*Stomolophus*	*meleagris*	Protocol 3	703	959,516	909,401	msats	591,608	-	92,668
MaF 9	Coral	*Porites*	*panamensis*	Protocol 1	702	3,449,711	3,298,155	msats	1,549,718	18,628 (50x)[5]	7.322
Total						10,838,641	10,233,942				

[1] Only includes high quality reads with inserts of 250 bases; excluded reads generally due to short insert length due to degraded input DNA.

[2] Identified using default parameters in PAL-finder (Castoe *et al*. 2012).

[3] Díaz-Jaimes *et al*. (2016)

[4] Galván-Tirado *et al*. (2016)

[5] Del Rio-Portilla *et al*. (2016).

Protocol 1: EZNA Tissue DNA KIT (Omega Bio-Tek, USA); Protocol 2: Aljanabi and Martínez (1997); Protocol 3: CTAB-Phenol/Chloroform.

Supplemental Figures

Figure S1. Adapterama overview. Adapterama is an integrated system of constructing libraries for next-generation DNA sequencing, whether by the iTru approach or iNext approach. The overall process for constructing dual-indexed Illumina libraries is depicted here. A variety of DNA inputs and application-specific processes are used to attach adapter stubs to the DNA of interest (black). In this illustration, we show a Y-yoke stub, but other configurations are possible and will be demonstrated in subsequent papers in the *Adapterama* series. Limited-cycle PCR is then used to extend the adapters, making them fully functional for sequencing on Illumina instruments and also adding dual indexes. The P5 (maroon) and P7 (light green) regions on the molecule are complementary to oligonucleotides present on Illumina flow cells, allowing for hybridization and clonal amplification. The i5 (dark green) and i7 (light blue) indexes can be used for multiplexing. The Read 1 (R1, violet) and Read 2 (R2, red) primer binding sites are complementary to the sequencing primers, enabling sequencing of the library molecules on the flow cell. The R1 and R2 primer binding sites also contain regions with identical sequence (shown in orange) that are used to facilitate the Y-yoke adapters. Thus, the full R1 and R2 sequences include the regions in orange (see Fig. 2). The color scheme matches Figure 1 and is used throughout, except as noted.

Figure S2. Overview of iNext library construction. Color schemes and naming conventions follow those of Figure 1 & S1. During library construction, sheared genomic DNA with C overhangs is ligated to stubby Y-yoke adapters with G overhangs (cf. Fig. 2). The C overhangs prevent chimeric ligation of genomic DNA molecules, and the G overhangs prevent ligation

among adapters (adapter dimers). Adapters are phosphorylated (indicated with a "P" at the 5' position), which allows ligation of stubs to genomic DNA. During limited cycle PCR, iNext5 and iNext7 primers anneal to the ends of the Y-yoke adapters to produce full-length, double-indexed molecules (cf. Figs. 2, S3, and S4).

Figure S3. Library preparation PCR and product formation. During PCR, adapter-ligated DNA molecules (i.e., the ligation products) react with index containing primers to create the double-indexed, full-length library molecules. During the first PCR cycle, only the iTru7 primer binds to the denatured strand at the 3' end of the template molecules. The iTru5 primer has the same sequence and orientation as the 5' ends of the molecule, and therefore cannot anneal during the first cycle. The product of the first cycle, an abbreviated single-indexed molecule, creates a sequence complementary to the iTru5 primer. During the second cycle, both primers are able to anneal to a denatured strand, creating a dual-indexed, truncated molecule with an overhang at the 3' end on the bottom strand. In the third cycle, a full-length, dual-indexed (i5 index on top strand, i7 on bottom) library-prepared molecule is made.

Figure S4. Comparison of iTru vs. iNext library molecules. Complete double-stranded library molecules are illustrated. The color schemes used previously have been simplified so that the Y-yoke portions of Read1 (R1) and Read2 (R2) are now simply shown as lighter colors of the non-complementary regions. Although functionally equivalent, the R1 and R2 regions of iTru and iNext have no sequence similarity. In contrast, the P5 and P7 regions are identical. Although the iTru and iNext indexing regions (i5 and i7) are illustrated in the same colors and

949 draw from the same pool of tags (Faircloth and Glenn, 2012), there is no correspondence in
950 numbering (i.e., iTru5_01_A index ≠ iNext5_01_A index).

951 **Figure S5. Edit distances between Illumina sequence tags**. Figure shows the sequence tags
952 used by Illumina on the P7 side of the library construct in the Illumina TruSeq HT kits (D###)
953 and the Illumina TruSeq (TS-##) kits. Note that several of the edit distances within and between
954 each set of sequence tags are ≤3 (e.g. D702 vs. D710; D705 vs. TS-15), which precludes the use
955 of edit-distance error correction algorithms to recover sequence tags containing a sequencing
956 error.

957

958 **Figure S6. Complete iTru library molecule and sequencing primers highlighting the**
959 **complementary regions from the Y-yoke adapter.** The color scheme here is similar to those
960 used in Figs. 1, 3, S1, S2, and S3. Reads work similarly for iNext.

961

962 **Figure S7. Illustration of Illumina sequencing reads from libraries with a single (i7) index.**
963 Illumina still supports libraries with a single index; the i7 index (i.e., Indexing Read1) is always
964 used in these instances. If libraries of this type are mixed with iTru, or any other dual-indexing
965 libraries, and both index sequencing reads are obtained from the pool, an i5 sequence will be
966 generated, but different strands and thus positions will be sequenced based on which instrument
967 (indexing read2 primer) is used. The i5 sequence obtained will be GTGTAGAT from NextSeq
968 and MiniSeq, whereas the sequence ACACTCTT is obtained from MiSeq and HiSeq ≤2500
969 instruments. HiSeq ≥3000 instruments initially generate the sequence GTGTAGAT, but that is
970 reverse complemented to ATCTACAC by Illumina software. Because all Nextera-type libraries
971 are dual indexed, there is no similar situation for Nextera or iNext libraries.

Figure S8. Mean threshold cycle (C_T), ΔC_T (average), and ΔC_T (baseline) for iTru7 primers (panels A-C) and iTru5 primers (panels D-F). The iTru7_401_07 primer performed poorly during the first qPCR test, and we retested this same primer from a new aliquot of oligos. The iTru7_401_07 primer performed normally during this second retest.

Figure S9. Number of reads sequenced from each of the *E. coli* iTru libraries relative to the mean number of reads sequenced from all *E. coli* libraries (dotted line) and the 95% confidence interval around this mean (dot-dashed line).

Figure S10. Comparison of aggregate read counts and the percentage of reads generated for each sample across the larger scale test projects using the iTru system to prepare Illumina-compatible libraries.

Figure S11. The percentage of reads generated for each combination of iTru5 and iTru7 from a study of 90 cichlid fish lineages. Data were generated from a partial, PE150, Illumina NextSeq High Output run, and the target for each sample was 1.0% of the total reads generated across the partial run (blue). The heat map shows deviations from the optimal percentage.

Figure S12. The percentage of reads generated for each combination of iTru5 and iTru7 from a study of 183 carangimorph fish lineages. Data were generated from a partial, PE150, Illumina NextSeq High Output run, and the target for each sample was 0.5% of the total reads

994 generated across the partial run (blue). The heat map shows deviations from the optimal
995 percentage.
996
997 **Figure S13. The percentage of reads generated for each combination of iTru5 and iTru7**
998 **from a study of 100 ant lineages.** Data were generated from one lane of PE125 sequencing on
999 an Illumina 2500, and the target for each sample was 1% of the total reads generated across the
1000 entire run (blue). The heat map shows deviations from the optimal percentage.
1001
1002 **Figure S14. The percentage of reads generated for each combination of Illumina D5##,**
1003 **iTru5, and iTru7 from a study of 203 bird lineages.** Data were generated from two lanes of
1004 PE150 sequencing on an Illumina 1500 in Rapid Run mode, and the target for each sample was
1005 1% of the total reads generated across the entire run (blue). The heat map shows deviations from
1006 the optimal percentage.
1007
1008 **Supplemental Files**
1009 Supplemental File 1. Supplemental methods detailing design and testing of iNext adapter stubs
1010 and iNext Primers.
1011 Supplemental File 2. Example sample sheet used when demultiplexing libraries sequenced on
1012 Illumina platforms.
1013 Supplemental File 3. Edit distances between Illumina sequence tags.
1014 Supplemental File 4. Excel workbook with iTru adapters and iTru7 and iTru5 primers, along
1015 with ordering details, and sample sheet preparation (demultiplexing) information.
1016 Supplemental File 5. Generic primer plate layout.

Supplemental File 6. Protocol for preparing 1.25 nmol iTru primer aliquots for use during library preparation.

Supplemental File 7. Protocol for preparing double-stranded iTru adapters for use during library preparation.

Supplemental File 8. Protocol for preparing an inexpensive substitute for AMPure.

Supplemental File 9. Protocol for preparing iTru libraries.

Supplemental File 10. Templates for calculating base diversity within sets of indexes.

Supplemental File 11. Templates to use when combining libraries needing different read counts into a single Illumina run. Templates are given on separate tabs for pooling based on the desired number of reads or the desired percentage of a run.

Supplemental File 12. qPCR results from iTru primer tests.

Supplemental File 13. Summary information from iTru and iNext *E. coli* libraries.

Supplemental File 14. Summary read counts from other projects using the iTru system to index libraries.

Supplemental File 15. Excel workbook with iNext stub adapters and iNext5 and iNext7 primers, along with ordering details, and sample sheet preparation (demultiplexing) information.

Supplemental File 16. Protocol for preparing iNext libraries.

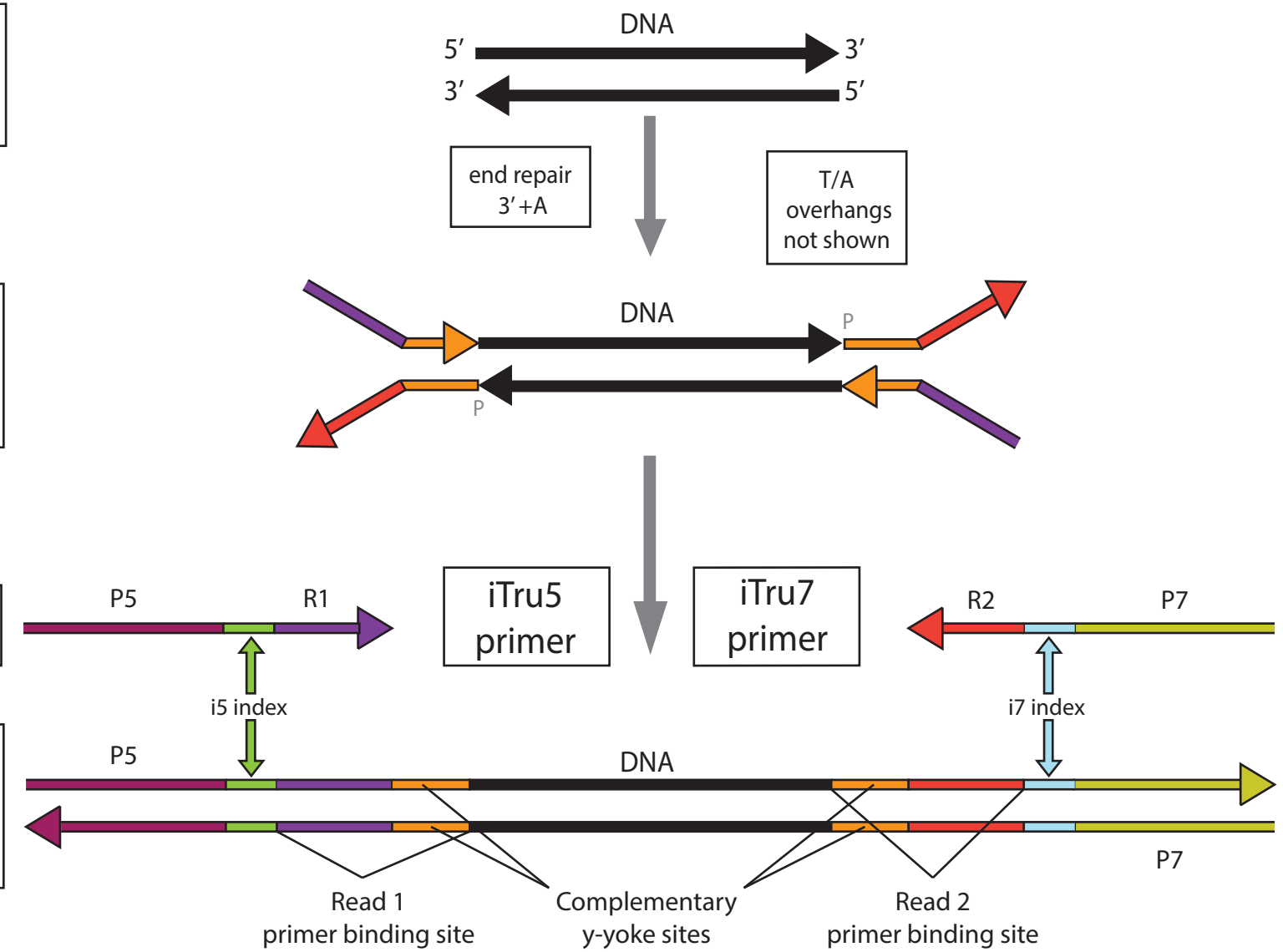
iTru Library Method

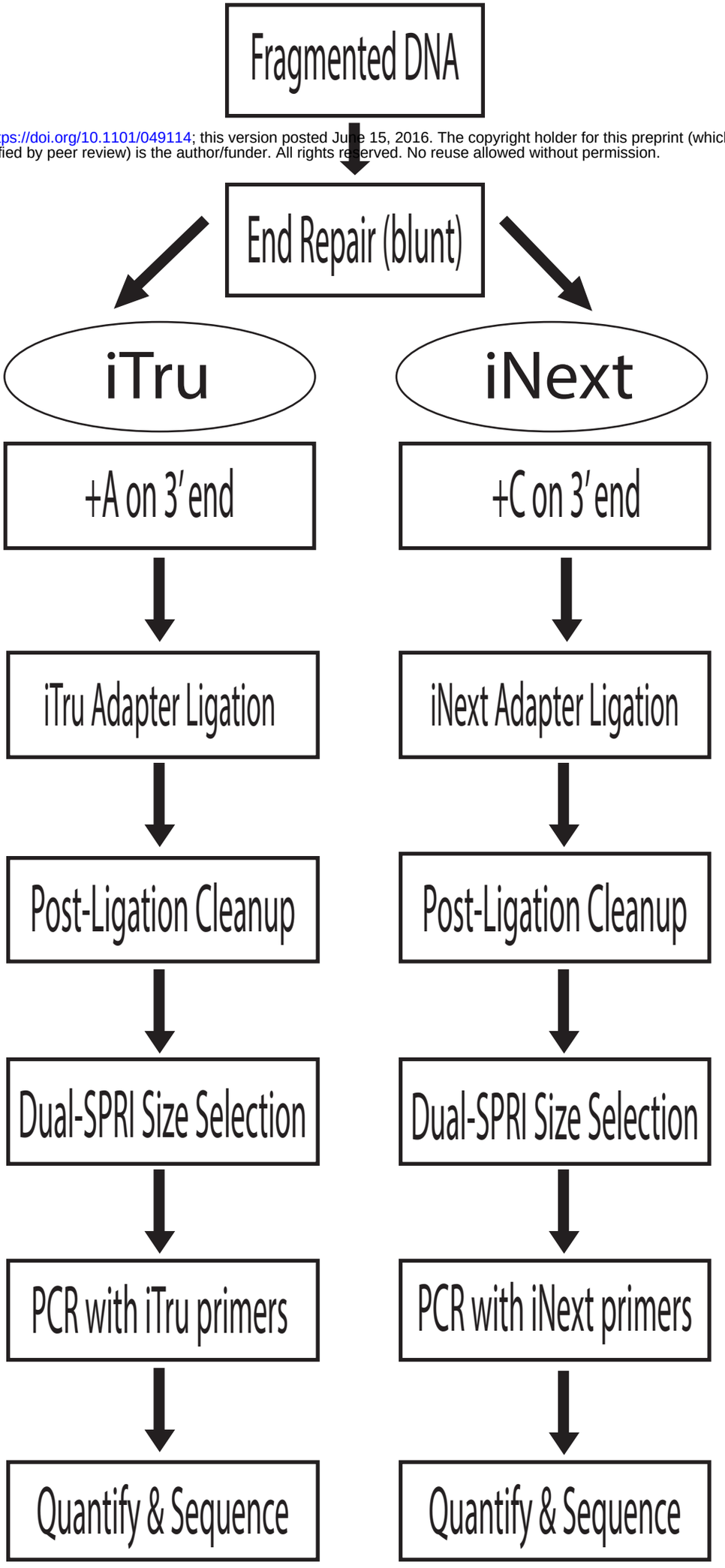
Fragmented DNA sample

Ligate stubby Y-yoke adapters

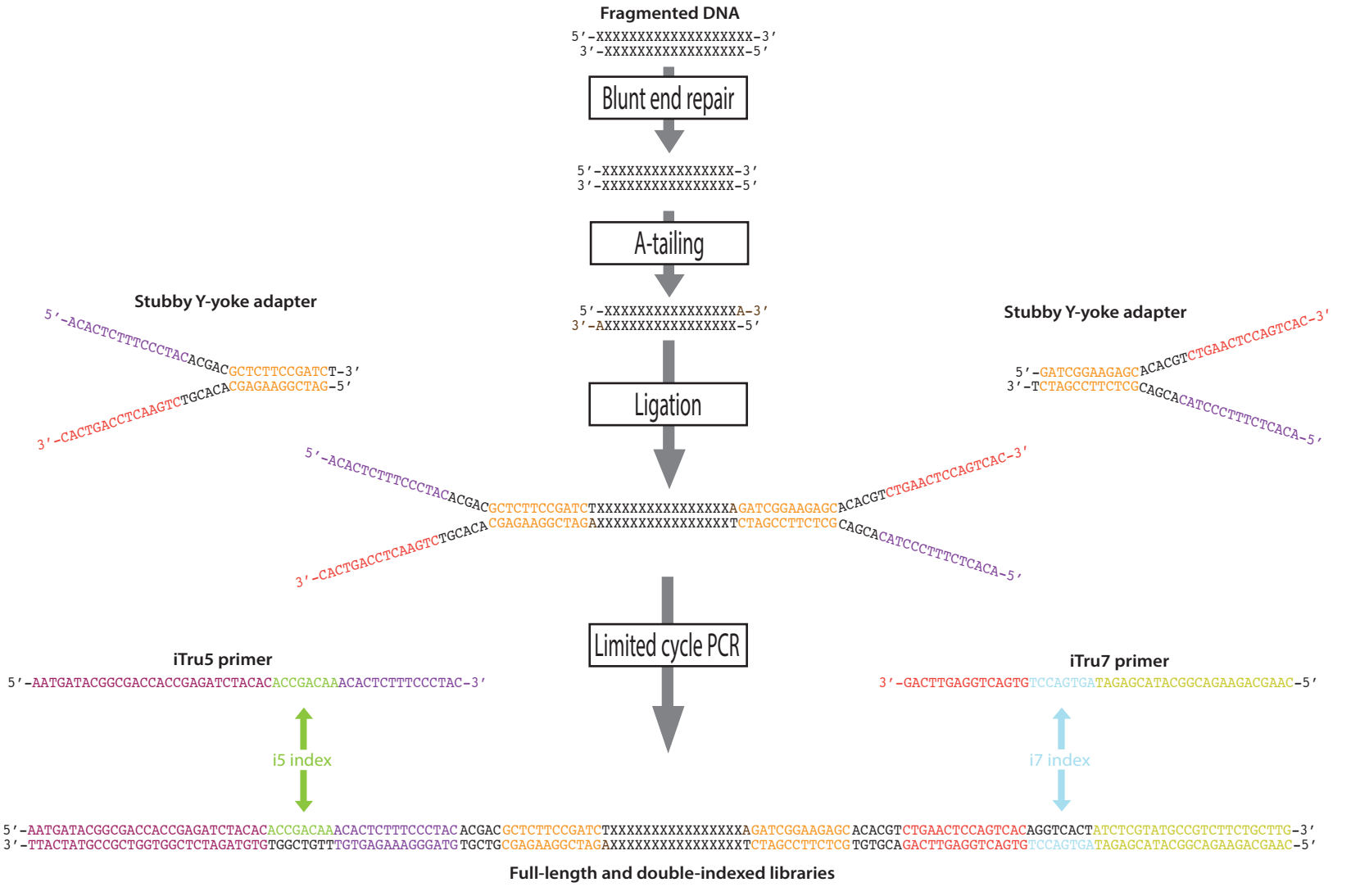
Limited cycle PCR

Full-length & double indexed





Detailed View of iTru Library Method



Sequencing Reads for Paired-End and Dual-Indexes