Parallel tagged sequencing on the 454 platform

Matthias Meyer, Udo Stenzel & Michael Hofreiter

Max Planck Institute for Evolutionary Anthropology, Department of Evolutionary Genetics, Deutscher Platz 6, D-04103 Leipzig, Germany. Correspondence should be addressed to M.M. (mmeyer@eva.mpg.de).

Published online 31 January 2008; doi:10.1038/nprot.2007.520

Parallel tagged sequencing (PTS) is a molecular barcoding method designed to adapt the recently developed high-throughput 454 parallel sequencing technology for use with multiple samples. Unlike other barcoding methods, PTS can be applied to any type of double-stranded DNA (dsDNA) sample, including shotgun DNA libraries and pools of PCR products, and requires no amplification or gel purification steps. The method relies on attaching sample-specific barcoding adapters, which include sequence tags and a restriction site, to blunt-end repaired DNA samples by ligation and strand-displacement. After pooling multiple barcoded samples, molecules without sequence tags are effectively excluded from sequencing by dephosphorylation and restriction digestion, and using the tag sequences, the source of each DNA sequence can be traced. This protocol allows for sequencing 300 or more complete mitochondrial genomes on a single 454 GS FLX run, or twenty-five 6-kb plasmid sequences on only one 16th plate region. Most of the reactions can be performed in a multichannel setup on 96-well reaction plates, allowing for processing up to several hundreds of samples in a few days.

© 2008 Nature Publishing Group http://www.nature.com/natureprotocols

INTRODUCTION

Rationale

Over the last three decades, Sanger sequencing[1] has been the dominant DNA sequencing technology in all areas of life sciences, used to retrieve individual sequences or to decipher entire genomes. Although the throughput of this technology has gradually increased over time, it has now been exceeded by recently developed next-generation sequencing technologies[2], such as 454 (Roche)[3], Solexa (Illumina)[4] and SOLiD (ABI). These technologies have increased the number of sequences obtained in a single run of a machine by several orders of magnitude, from mere hundreds to hundreds of thousands or even millions. Their superior efficiency in terms of both cost and time per sequenced nucleotide has not only spawned exploration in new sequencing fields, for example, ultra deep amplicon sequencing[5] or paleogenomics[6], but has also replaced Sanger sequencing in some of its ancestral domains, such as genome sequencing[3,7] and serial analysis of gene expression[8].

Among the next-generation sequencing technologies, 454 currently offers by far the highest read length, which is ~250 bp on the GS FLX platform, not far from the 700 bp achieved through routine Sanger sequencing. However, despite its comparatively low throughput, Sanger sequencing is still used for many everyday applications, for example, amplicon sequencing and the sequencing of DNA fragments, a few kilobases long by primer walking. One important reason for this lies in a conceptual difference between Sanger and 454 sequencing, which affects the number of samples that can be processed in parallel. Whereas in Sanger sequencing each sequence read is derived from a separate sequencing reaction, 454 uses emulsion PCR[9] to amplify a pool of templates in a single reaction vessel before sequencing. Within one emulsion PCR, no information is retained about a sequence's sample origin. Thus, to process several samples in parallel, these must be kept in separate pools, physically subdivided from each other not only during library preparation, but also during bead-emulsion-amplification and sequencing. However, the 454 sequencing plate can only be divided into a maximum of 16 regions, each of which yields on average 3 Mb of sequence. In many cases, this amount of sequence data produces an unnecessarily high redundancy in coverage. For example, if a 6-kb plasmid is shotgun sequenced on one 16th GS FLX plate region, it will be covered 500-fold on average. If it were possible to retain information about the sample origin of the obtained sequence reads, the same capacity could be used to sequence 25 such plasmids to 20-fold coverage.

The method described here, called parallel tagged sequencing (PTS), allows for parallel sequencing large numbers of double-stranded DNA (dsDNA) samples on the 454 platform[10]. This is achieved by barcoding each sample with a specific sequence tag. After pooling the tagged DNA samples, library preparation and sequencing, the tag sequences are used to identify each sequence's sample origin. The protocol (illustrated in **Fig. 1**) begins by blunt-end repairing each sample in separate reactions. Subsequently, barcoding adapters are ligated to both ends of the molecules. These adapters comprise single self-hybridized oligos containing a sequence tag and an *Srf*I restriction site. After ligation, the resulting single-strand nicks are removed by fill-in using a strand displacing polymerase. The barcoded samples are then quantified and pooled in ratios reflecting the desired relative sequence representation. After dephosphorylation, half of the adapter is cut off using *Srf*I[11], a rare cutting restriction enzyme with restriction sites approximately every 150 kb in the human genome. *Srf*I leaves 5′ phosphates for the ligation of universal 454 adapters during sequencing library preparation. The dephosphorylation step excludes unreacted molecule ends from sequencing.

PTS offers several important features providing both efficient use of sequencing resources and high data reliability. First, all reactions are completed with ~100% efficiency, ensuring highly homogeneous sequence representation among samples. Second, background sequences without a sequence tag are efficiently excluded from sequencing by dephosphorylation in conjunction with the use of a restriction enzyme. The *Srf*I restriction site produces a run of Gs before the tag and immediately adjacent to the key sequence used by the 454 system for quality controls. As the last nucleotide of this key is also Gua, all nucleotides remaining from the *Srf*I restriction site are inserted within a single flow cycle without significantly reducing the read length of the following

Illumina Exhibit 1014, Page 1 of 12

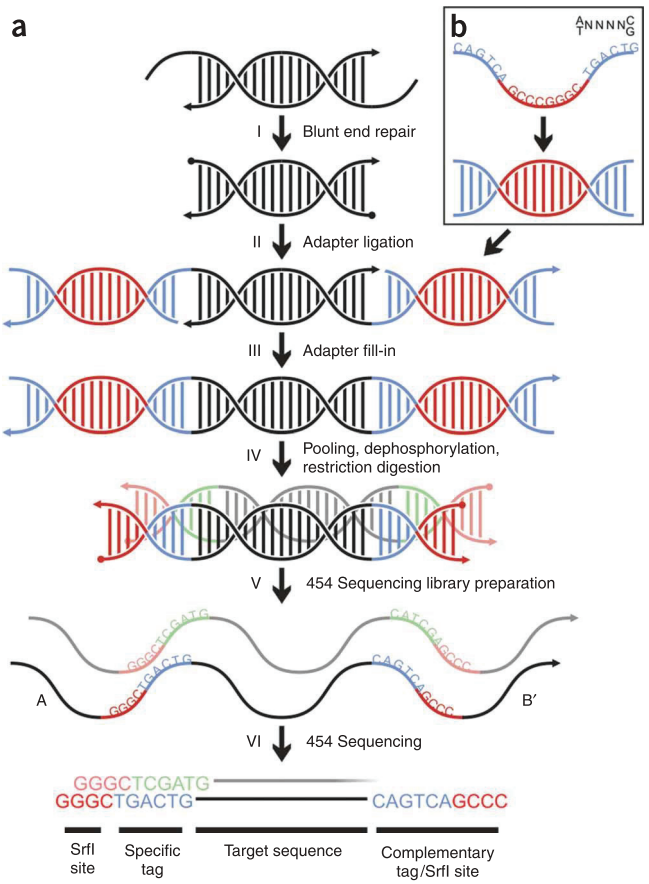

Figure 1 | Overview of the tagging protocol. (**a**) Each DNA sample is blunt-end repaired (I, Steps 2–8), before sample-specific barcoding adapters are ligated to both ends of the molecules (II, Steps 9–13). Nicks resulting from the ligation are removed by strand displacement with Bst polymerase (III, Steps 14–16). The barcoded samples are pooled in equimolar ratios and unligated molecule ends are excluded from sequencing through dephosphorylation and restriction digestion (IV, Steps 18–25). A single-stranded 454 sequencing library is prepared from the sample pool; this includes the blunt-end ligation of universal 454 adapters to the template molecules and isolation of correctly ligated molecules as single strands (V, Step 26). After sequencing (VI, Steps 27 and 28) the sequence reads are sorted according to their tag sequences (Step 29). Before downstream data processing, the sequence tags are removed from the 5′ ends and, if applicable, the 3′ ends of the reads. (**b**) Barcoding adapters comprise single self-hybridized palindromic oligonucleotides, carrying an *Srf*I restriction site in the middle (GCCCGGGC), a sequence tag at the 3′ end and the reverse complementary tag sequence at the 5′ end. Each sequence tag may start with either an A or T, followed by several freely chosen nt, and ends in C or G. No homopolymers are allowed within the tag sequence.

sequence. Third, the tag design is particularly robust to sequencing errors in and around homopolymers, which are known to be the most common errors in 454 sequencing. In our experience, ~97% of the sequences can be assigned to their sample origin with an extremely low false-assignment rate. Finally, the protocol is optimized for reaction setup using multichannel pipettes and a 96-well plate format, minimizing the time required for setup.

Applications of PTS

In existing 454 applications, physical subdivision of the 454 sequencing plate is frequently used to process up to 16 different samples in one run. As this requires covering the sequencing plate with a gasket, the overall number of sequences retrieved from one run is reduced by half (**Table 1**). Using PTS instead of physical subdivision for such applications immediately doubles the sequencing throughput. Moreover, as theoretically an unlimited number of tags can be produced, PTS overcomes any limitation on the number of samples that can be processed in parallel. In principle, PTS can be applied to all types of double-stranded nucleic acid samples, allowing an efficient switch from Sanger to 454 sequencing. We are currently using PTS particularly for two sequencing applications, which are discussed below.

Shotgun sequencing contiguous DNA segments. Owing to its high throughput and the absence of microbial subcloning, the 454 technology enables faster and cheaper shotgun sequencing compared to the Sanger methodology. Using PTS, this power can be fully exploited for parallel sequencing contiguous DNA segments

© 2008 Nature Publishing Group http://www.nature.com/natureprotocols

TABLE 1 | Sequencing throughput of the GS 20 and GS FLX platforms. 454 Sequencing plates can physically be subdivided into a minimum of two and a maximum of 16 regions. When using 16 plate regions, roughly half of the output is lost, as parts of the plate are covered with a gasket. In contrast, PTS allows for sequencing hundreds of samples in parallel without requiring physical subdivision, thereby retaining the maximum throughput.

Sequencing platform		**GS 20**			**GS FLX**	
Average read length		~100 bp			~250 bp	
Plate region	**1/16th**	**1/4th**	**1/2**	**1/16th**	**1/4th**	**1/2**
Reads per region	6,300	33,000	100,000	12,000	70,000	210,000
Base pairs per regions	630 kb	3.3 Mb	10 Mb	2.88 Mb	16.8 Mb	50.4 Mb
Base pairs per plate	10 Mb	13.2 Mb	20 Mb	46 Mb	67 Mb	101 Mb
Number of samples per plate region that can be processed in parallel using PTS						
17 kb segments, for example, mtDNA genomes (shotgun sequenced, average 20-fold coverage)						
	2	10	29	8	49	148
6 kb segments, for example, plasmids (shotgun sequenced, average 20-fold coverage)						
	5	28	83	24	140	420
PCR products (<100/250 bp length, average 40-fold coverage)						
	158	825	2,500	300	1,750	5,250

Illumina Exhibit 1014, Page 2 of 12

a

	Sequences	Number	% of total	Average coverage	Fully assembled
	Total	14,780			
Tag 1	TGACGTG	683	4.6	9.6	Yes
Tag 3	AGCACTG	938	6.3	13.0	Yes
Tag 5	TCTGATG	1,058	7.2	14.6	Yes
Tag 7	ACTGTCG	1,091	7.4	15.0	Yes
Tag 9	TGTCTCG	1,019	6.9	14.1	Yes
Tag 11	TCATGCG	1,107	7.5	15.5	Yes
Tag 13	ACGAGCG	776	5.3	12.6	No
Tag 15	TGAGACG	901	6.1	12.5	Yes
Tag 17	ACACACG	987	6.7	13.8	Yes
Tag 19	TCACTAG	888	6.0	12.5	Yes
Tag 21	ATGTGAG	1,027	6.9	14.1	Yes
Tag 23	TGCAGAG	933	6.3	12.9	Yes
Tag 25	TGTGCAG	1,525	10.3	20.3	Yes
Tag 27	TGCTCTC	1,284	8.7	17.6	Yes
	No tag	563	3.8		

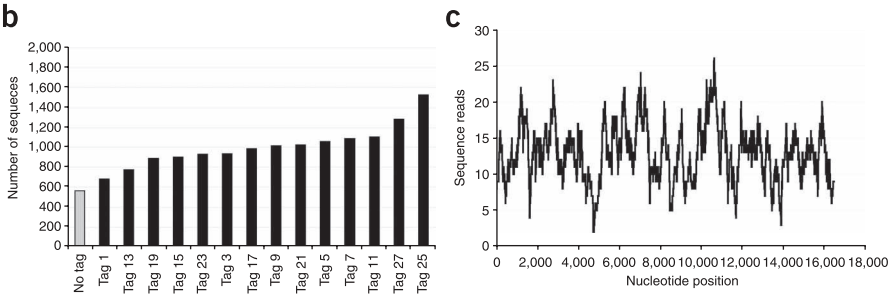

Figure 2 | PTS of 14 complete human mtDNA genomes (~16.5 kb) on a small GS FLX plate region to on average 14-fold coverage. The mtDNA genomes were amplified in two overlapping long-range PCRs as described previously[10]. The long-range PCR products were then quantified, pooled in equimolar ratios and nebulized. Between 100 and 200 ng of each sample were used as templates for the tagging reactions using barcoding adapters with 7-bp sequence tags differing by at least three substitutions. From 40 barcoding adapters that had been synthesized and diluted in a single batch in order from 1 to 40, no immediate neighbors were used to obtain full power for detecting cross-contamination among barcoding adapters. After barcoding, the samples were pooled in equal mass ratios. (**a**) Table of sequencing results showing the number of sequences and coverage obtained for each sample. As a result of inaccuracies in quantification or pipetting when pooling the long-range PCR products, one sample exhibited uncovered positions. These could be filled-in by deeper sequencing or single Sanger reads. No sequences with sequence tags from unused barcoding adapters were observed, indicating no detectable presence of cross-contamination among the barcoding adapters and the absence of sequencing errors potentially leading to false assignment of sequences to their sample origin. Thus, the best estimate of the false-assignment frequency in this experiment is zero. (**b**) Bar chart visualizing the sequence representation among samples, ordered from lowest to highest. (**c**) Exemplary coverage plot for one of the mitochondrial genomes (tag 23).

© 2008 Nature Publishing Group http://www.nature.com/natureprotocols

from numerous samples, such as plasmids or target regions pre-amplified by long-range PCR (see ref. 10, **Fig. 2**). For example, when sequencing is performed on the new 454 GS FLX platform, up to 300 complete mitochondrial genomes or a comparable number of nuclear DNA fragments of similar length (~17 kb) can theoretically be sequenced to 20-fold coverage in parallel in a single run (see **Table 1**). In this way, population data produced by re-sequencing can be obtained with unprecedented speed. In contrast to Sanger-based primer walking approaches, shotgun sequencing long-range PCR products does not require *a priori* sequence information for designing sequencing primers and saves time and costs for setting up individual PCRs and sequencing reactions.

Sequencing pooled amplicons. This application is useful when short sequences within the 454 read length limit are desired, as in population studies using ancient DNA or DNA from museum specimens. As 454 sequence reads stem from single template molecules, miscoding base damage and contamination can be readily identified without microbial subcloning of the PCR products. This allows for cost- and time-efficient sequencing of pooled amplicons from multiple samples, while retaining the highest standards in ancient DNA and museum research[12,13]. Phylogenetic and population genetic studies are also increasingly performed using multiple short nuclear sequences, totaling from a few to 30 kb of sequence[14–16]. In such applications, complete data sets for whole species groups as well as large population samples could be obtained in either a single or partial 454 run. In addition to low coverage sequencing of many pooled amplicons, PTS can be used for deep sequencing fewer amplicons in parallel.

Comparison to other barcoding methods

Although other methods for barcoding and sample multiplexing on the 454 platform have been previously introduced, these methods are limited to the parallel sequencing of PCR products. In contrast, PTS is the first barcoding method to allow for parallel sequencing all types of dsDNA samples. It is also the only currently available method for barcoding shotgun DNA libraries and pooled PCR products.

The previously reported barcoding methods used 5′-tagged PCR primers to distinguish PCR products derived from different sources[17–19]. The universal 454 adapter sequences were either included as additional 5′-tails or added in the regular 454 library preparation process. This approach is simple, quick and efficient for sequencing short (<250 bp) homologous PCR products from different samples, because combinations of tagged forward and reverse primers can be used to barcode a large number of samples. However, when dealing with many different targets, this approach becomes cost-prohibitive and prone to confusion, because sets of primers must be synthesized for each target under study, and the primers must be added separately both to each PCR and the corresponding control reaction. PTS is preferable in this case, as it is suitable for simultaneously barcoding a pool of PCR products. It does not require changes to the experimental design of existing PCR applications and provides the flexibility of choosing the sequencing strategy after amplification. Furthermore, because they consist only of single short self-hybridized oligos, barcoding adapters for PTS are cheap to synthesize and can be reused in subsequent experiments.

Another method has recently been introduced for parallel sequencing small RNAs[20]. It involves stepwise single-stranded ligation of universal adapters to both ends of the RNA molecules. Barcoding is then achieved by re-amplification with tagged PCR primers. Although this method may be suitable for barcoding small RNAs, the protocol is very complex and has yet to be applied to dsDNA samples.

Limitations of the method

One limitation of the method arises through the use of a restriction enzyme. With its GC-rich 8 bp recognition sequence, *Srf*I is a rare cutter in mammalian genomes, with restriction sites approximately only every 150 kb in the human genome. However, it may cut more

Illumina Exhibit 1014, Page 3 of 12

often in GC-rich bacterial genomes. As the 454 universal adapters are added after the *Srf*I restriction step, only sequence coverage immediately around an *Srf*I site is lost, and gaps can be filled with single Sanger sequence reads. If sequencing templates are known to contain *Srf*I sites, it is possible to enzymatically methylate all *Srf*I sites before adapter ligation using CpG methyltransferase according to the supplier's protocol (http://www.neb.com). Owing to the inability of *Srf*I to cut CpG methylated restriction sites, this should effectively mask all restriction sites, thereby eliminating restriction occurring within template molecules.

Another important issue is the quality of the resulting sequences. With regard to substitutional errors, well above 99.99% accuracy was consistently reported for shotgun consensus sequences on the GS 20 platform[3,10,21,22], and 99.92% were estimated for single reads in a recent study[23]. However, single base pair insertions and deletions (indels) occur with considerable frequency both within and around homopolymer regions, often persisting even at high coverage. In shotgun consensus sequences of human mitochondrial genomes[10], we recently observed indel errors at a frequency of 0.27%, although previous estimates from shotgun consensus sequences were about ten times lower[3,21,22]. However, all current estimates should be considered with caution, as the error rate varies among different versions of the 454 assembly programs newbler and runMapper (see ref. 21, and M.M., unpublished data), and may also differ among runs. As single base pair indels can be identified as frame-shift mutations in coding sequences or by comparison to closely related sequences, they usually have no practical impact on sequence usability.

Experimental design

Several points should be considered before large-scale sequencing projects are performed using PTS.

Sample requirements. In principle, every double-stranded nucleic acid sample with natural 5′-ends (hydroxyl or phosphate) is a suitable template for parallel sequencing using PTS. However, there are upper and lower limits on template size. The upper limit is defined by the 454 process, as the maximum read lengths obtained on the GS 20 and GS FLX platforms are ~100 and 250 bp, respectively. In addition, fragments of 800 bp or more amplify poorly in emulsion PCR. Adequate fragment length distributions, for example, for shotgun sequencing, can be achieved by DNA shearing. The lower size limit is introduced through the SPRI bead purification steps in the PTS protocol. SPRI bead purification[24] is quick and efficient, but does not recover molecules <80–100 bp. If shorter molecules need to be sequenced (50–100 bp), all SPRI purification steps can be replaced by MinElute Spin column purification (Qiagen), using the same elution volumes and buffers without additional changes.

The minimal material requirements for PTS are very low. As 454 sequencing is possible from picogram amounts of DNA[25], and there are no significant losses in the tagging protocol, <1 ng of initial material per sample is theoretically sufficient. However, the sequence representation of each barcoded sample depends on its relative concentration in the sample pool, and is thus affected by the accuracy of DNA quantification. For optimal results, we recommend measuring DNA concentration after barcoding (Step 18). This strategy requires at least 100 ng of initial material for PicoGreen quantification. If less material is available, DNA amounts can be measured and adjusted before barcoding (after Step 1). Although requiring very little material, the latter strategy yields less homogeneous sequence representation, as handling variation may cause different recoveries during the purification steps. For shotgun sequencing long-range PCR products, we generally recommend using long-range PCR kits from Roche (Expand Long-Range dNTPack, Expand 20kbPlus PCR Systems), which in our hands yield superior results as compared with other suppliers.

Choosing a tag length. To avoid falsely assigning sequences to their respective samples, not all possible tags of a certain length should be used, as single substitutions attributable to sequencing errors could convert one tag into another. A tag length of 6 nt produces only 72 different tags that are at least two substitutions apart, and 21 different tags that are at least three substitutions apart. These numbers are 173 and 52 for 7-nt tags, and 475 and 130 for 8-nt tags, respectively. With a minimal distance of two substitutions between 6-nt tags, we previously estimated a false-assignment rate of ~0.35% on the GS 20 platform[10]. Using the new GS FLX platform and 7-nt tags, this number drops to 0.03% at a minimal distance of two, and <0.01% at a minimal distance of three substitutions, respectively (M.M., unpublished data). However, these numbers should be considered as rough estimates only, as they may vary among runs and with the purity of oligos. We recommend independently estimating the false-assignment rate in each experiment (see **Box 1** and QUALITY CONTROL, below). Since on the GS FLX platform the read length has increased to ~250 bp compared with 100 bp on the previous GS 20 system, tag lengths of 7 or 8 nt do not significantly reduce the amount of usable sequence data obtained by PTS, but provide the opportunity to sequence hundreds of samples in parallel at extremely low false-assignment rates.

Coverage requirements and sequencing strategy. An issue that requires careful evaluation before PTS is begun is the amount of sequence coverage required. For shotgun sequencing, in our experience 10- to 20-fold average coverage is sufficient for re-sequencing mono-allelic sequences, such as mitochondrial genomes. Indel sequencing errors around homopolymers can be eliminated by comparison to a reference sequence. If no closely related sequences are available for comparison, higher sequence coverage (~30-fold) is preferred for obtaining sequences with a low indel rate. If nuclear sequences with two potential alleles are sequenced, higher coverage is necessary to ensure that both alleles are detected in heterozygous samples. In general, the coverage requirements must be chosen according to the specific needs of a study.

While estimating the coverage requirements for a study, it is important to understand that sequence representation is approximately normally distributed among samples. Whereas most samples will be covered by the desired number of sequence reads, a few samples will be covered higher or lower. For samples quantified after barcoding (see SAMPLE REQUIREMENTS, above), we usually observed at maximum ~50% deviation from the mean coverage (**Fig. 2** and M.M., unpublished data). This can be compensated for with either higher coverage sequencing *a priori*, or by subsequently filling in sequences from under-represented samples in an additional run on a small plate region. For large-scale projects, where one or several full runs will be completed, the sequencing resources can be optimally exploited by initially sequencing part of the sample pool on only a small plate

© 2008 Nature Publishing Group http://www.nature.com/natureprotocols

Illumina Exhibit 1014, Page 4 of 12

BOX 1 | ESTIMATING THE FALSE-ASSIGNMENT FREQUENCY

The reliability of PTS should be independently evaluated in each experiment. This can be achieved by estimating the false-assignment frequency, that is, the frequency at which false assignment of sequences to their sample origin is expected, based on the occurrence of sequence reads that carry tag sequences from unused barcoding adapters. The ability to detect false assignment improves as the number of barcoding adapters that remain unused in an experiment increases.

$$\text{False-assignment frequency} = \frac{F}{T} \times \frac{N}{A - N}$$

F, Number of sequences carrying tags from unused barcoding adapters
T, Total number of sequence reads obtained in the experiment
N, Total number of barcoded samples that were sequenced in parallel
A, Total number of barcoding adapters within the chosen category that have actually been synthesized (e.g., 52 if all 7-bp tags differing by at least three substitutions were synthesized)
The formula is based on the assumption that all tags can be converted into one another with the same probability. It is a composite estimate of false assignment that occurs due to sequencing errors and cross-contamination of barcoding adapters during synthesis and dilution. It does not consider the possibility that cross-contamination is introduced while preparing samples for PTS or setting up the blunt-end repair and ligation reactions (Steps 1–10) and is unlikely to detect punctual contamination. Thus, careful pipetting is strongly advised.

region. Subsequently, samples can be re-pooled according to the observed sequence representation, thereby guaranteeing optimal sequence representation among samples during the large-scale sequencing phase.

Quality control. Finally, as the assignment of tags to the correct sample source is critical, some quality control should be performed. The two major factors leading to false assignment of sequences are cross-contamination of barcoding adapters and sequencing errors. By testing a subset of barcoding adapters in a small-scale experiment before large-scale adoption of PTS, it is possible to monitor whether cross-contamination of adapters occurred during synthesis or dilution. If cross-contamination goes undetected, misassignment of sequences will occur. Moreover, in such a small-scale preliminary experiment, the occurrence of tag sequences from unused adapters can be monitored and used to estimate the false-assignment rate due to sequencing errors and/or errors occurring during oligo synthesis. By randomly omitting a small subset of barcoding adapters used in a study, the same quality control is advised for every experiment using PTS (**Box 1**). As all molecules carry sequence tags on both ends, another independent, albeit less stringent, quality check can be performed by comparing the 5′ and 3′ tag sequences in reads where the ends of molecules have been reached. The repeated occurrence of identical false tag pairs indicates that cross-contamination persists among barcoding adapters or was introduced during pipetting. It should be noted, however, that the identification of 3′ tag sequences is less reliable due to the higher sequencing error rate near the ends of reads and possible misidentification of 3′ adapter sequence starting points. In addition, for many samples with relatively long mean fragment sizes, such as shotgun libraries, the majority of sequences will terminate before the end of the molecule and the 3′ adapter are reached.

When combining PTS with physical separation of the sequencing plate, avoid sequencing different libraries containing the same sequence tags on neighboring regions, as occasionally leakage of sequencing beads occurs.

© 2008 Nature Publishing Group http://www.nature.com/natureprotocols

MATERIALS

REAGENTS

- T4 DNA polymerase (Fermentas, cat. no. EP0062)
- T4 polynucleotide kinase (Fermentas, cat. no. EK0032)
- T4 ligase (Fermentas, cat. no. EL0331), including 50% PEG-4000 solution and 10× ligation buffer
- Bst DNA polymerase, large fragment (NEB, cat. no. M0275S), including 10× ThermoPol buffer
- Calf-intestine phosphatase (NEB, cat. no. M0290S), including 10× NEBuffer 3
- *Srf*I restriction enzyme (Stratagene, cat. no. 501064), including 10× universal buffer
- 10× Buffer Tango (Fermentas, cat. no. BY5)
- ATP (Fermentas, cat. no. R0441), 100 mM stock solution
- dNTPs (GE Healthcare, cat. no. US77119-500UL), 25 mM each
- BSA (Sigma-Aldrich, cat. no. B4287), powder for preparation of a 10 mg ml^{-1} stock solution in water
- Water, HPLC-grade (Sigma, cat. no. 270733)
- Absolute ethanol (Merck, cat. no. 1.00983.2500)
- TE buffer (many suppliers or self-made); 10 mM Tris–HCl, 0.1 mM EDTA, pH 8.0
- EB buffer (supplied with MinElute PCR Purification kit); 10 mM Tris–HCl, pH 8.5
- DNA-loading dye (Fermentas, cat. no. R0611)
- Ethidium bromide (Sigma, cat. no. 46067), 1% solution ! CAUTION Mutagen and potential carcinogen.
- TBE electrophoresis buffer (Sigma, cat. no. 51309), 10× concentrate
- MinElute PCR purification kit (Qiagen, cat. no. 28006)
- PicoGreen dsDNA quantitation reagent (Invitrogen, cat. no. P11495)
- Oligonucleotides (Metabion), desalted, lyophilized. Sequences for sets of barcoding oligos with 6–8-nt tags are available in **Supplementary Table 1** online (see also REAGENT SETUP) ▲ CRITICAL Basic post-synthesis purification (desalting) is sufficient. Additional purifications, such as HPLC or PAGE, increase the risk of cross-contaminating oligonucleotides. The oligos should be synthesized on larger scales to suffice for several rounds of PTS.
- AMPure SPRI PCR purification kit (Agencourt, cat. no. 000130)
- GS DNA Library Preparation Kit (Roche, cat. no. 04852265001), including nebulizers and nebulization buffer ▲ CRITICAL Only 10 nebulizers and 20 ml nebulization buffer are supplied. Additional nebulizers can be obtained from Graham-Field (cat. no. BF61402). Nebulization buffer consists of 53.1% glycerol, 37 mM Tris–HCl, 5.5 mM EDTA, pH 7.5 (ref. 3).

EQUIPMENT

- 96-Well PCR plates
- Multichannel reagent basin (e.g., Thermo Scientific, cat. no. 9510027)
- Filter tips
- SPRIPlate 96R—Ring Magnet Plate (Agencourt, cat. no. 000219)

Illumina Exhibit 1014, Page 5 of 12

- Agarose gel electrophoresis unit
- NanoDrop spectrophotometer (NanoDrop Technologies)
- Stratagene MX 3005P QPCR System or any other qPCR system or microplate reader suitable for PicoGreen fluorescence measurements
- GS 20/GS FLX Genome Sequencer and associated equipment
- Software for data analysis (available with accompanying usage instructions at http://bioinf.eva.mpg.de/pts/)

REAGENT SETUP

Preparing barcoding adapters Barcoding adapters comprise of single self-hybridized oligos. Dissolve the oligos in TE to obtain 500 μM stock solutions. Create barcoding adapters by setting up separate reactions in PCR tubes containing in final concentrations 400 μM of one oligo and 1× T4 ligase buffer. Incubate in a thermal cycler with a temperature profile of 95 °C for 10 s and a ramp to 25 °C at a rate of 0.1 °C s^{-1}. Immediately freeze the barcoding adapters at −20 °C until further usage ▲ CRITICAL The barcoding adapters may be prepared and stored in PCR strip tubes or plates to allow for subsequent handling with multichannel pipettes. However, be extremely careful not to cross-contaminate the adapters during preparation or later handling. Preparation of aliquots may be useful for minimizing the number of handling cycles.

Generating a positive control template To control the efficiency of the tagging reactions, we strongly recommend carrying a PCR product as a control template alongside the samples for sequencing. Any PCR product producing a single distinct band of a size between 100 and 200 bp is suitable, as long as it is generated with *Taq* polymerase and unmodified PCR primers. Purify the PCR product using a MinElute spin column and adjust it to a concentration of 25 ng μl^{-1}. As 24 μl is needed for one control experiment, several PCRs may need to be pooled to obtain sufficient control template for several reactions.

Multichannel reaction setup Most steps in this protocol (Steps 2–18) can be performed using a multichannel pipette. After transferring liquids from a reagent basin to a 96-well PCR plate, mix the components by pipetting up and down or flicking the plate with your finger. Do not vortex enzyme-containing reactions before incubation. Seal the plate wells and quickly spin the plate in a centrifuge (∼1,000*g*) to collect the liquid at the bottom of the wells ▲ CRITICAL Multichannel pipetting requires extreme care to avoid sample cross-contamination by spilling liquids over the plate. We recommend using filter tips.

PicoGreen plate reads Detailed protocols for quantifying DNA with PicoGreen are supplied with the PicoGreen reagent or by the manufacturers of qPCR systems and microplate readers. Protocols can also be found in the literature[26,27].

© 2008 Nature Publishing Group http://www.nature.com/natureprotocols

PROCEDURE

Preparation of samples for tagging ● TIMING One or several days (depending on number and type of samples)

1| To be suitable for PTS, dsDNA samples must be purified from salts and enzymes that interfere with downstream reactions and concentrated to a maximum volume of 20 μl. EB and TE buffers are compatible with this protocol and suitable for elution and storage of DNA. Samples can be prepared using option A for sequencing short PCR products or option B for shotgun sequencing long-range PCR products, plasmids or genomic DNA.

(A) Purifying and pooling PCR products

(i) Purify PCR products using the MinElute PCR purification kit or SPRI beads according to the manufacturer's instructions. Elute in 20 μl TE.

(ii) Measure DNA concentrations in the purified PCR products using the nanodrop spectrophotometer or a PicoGreen quantification assay.

(iii) If applicable, pool PCR products that should be barcoded with the same sequence tag in molar amounts reflecting the proportion of sequences desired from each product. For example, to retrieve the same number of sequences from a 100-bp and a 200-bp PCR product, they should be pooled in a 1:2 mass ratio. Each pool should contain a final DNA amount of at least 100 ng.

(iv) A maximum volume of 20 μl can be used per tagging reaction. If 20 μl of a pool contain insufficient DNA amounts, concentrate the pool through MinElute purification and elution in 20 μl TE.

(B) Preparing shotgun DNA libraries

(i) Nebulize plasmid DNA, genomic DNA or purified long-range PCR products using nebulizers and reagents from the 454 library preparation kit following the manufacturer's instructions. Purify the nebulized DNA using one MinElute spin column and elute in 20 μl TE. At least 200 ng of sample should be used for nebulization, as up to 50% of the material is lost in this step. If desired, long-range PCR products derived from the same source may be pooled in equimolar ratios before nebulization.

(ii) Measure DNA concentrations in the purified PCR products using the nanodrop spectrophotometer or a PicoGreen assay.
▲ CRITICAL STEP When pooling long-range PCR products before nebulization, take care to measure their concentrations precisely. Inaccuracies in quantification will result in inhomogeneous sequence coverage between the long-range PCR products.
■ PAUSE POINT DNA samples diluted in TE buffer can be stored at −20 °C for several weeks.

Prepare multichannel reaction setup ● TIMING 1–3 h (for a full 96-well reaction plate)

2| A maximum of 94 samples can be processed on one reaction plate. Transfer 100–500 ng of each sample to a 96-well PCR plate. Add water to obtain a final volume of 20 μl. Add 20 μl of positive control sample (see REAGENT SETUP) to an empty well and reserve one empty well for a purification control (see Step 9).
▲ CRITICAL STEP Do not exceed 500 ng per sample. Higher amounts may facilitate chimera formation during the subsequent ligation step.

Illumina Exhibit 1014, Page 6 of 12

Blunt-end repair ● TIMING 1 h

3| Prepare a master mix for the required number of reactions. Mix carefully by pipetting up and down or flicking the tube with a finger. Do not vortex after adding enzymes.

Reagent	Volume (μl) per sample	Final concentration in reaction
Water	6.6	
Buffer Tango (10×)	4	1×
BSA (10 mg ml^{-1})	1	0.25 mg ml^{-1}
dNTPs (2.5 mM each)	1.6	100 μM each
ATP (10 mM)	4	1 mM
T4 polynucleotide kinase (10 U μl^{-1})	2	0.5 U μl^{-1}
T4 polymerase (5 U μl^{-1})	0.8	0.1 U μl^{-1}

▲ **CRITICAL STEP** Keep the master mix on ice if not immediately used to maintain full enzyme activity.

4| Add 20 μl of master mix to each sample to obtain a total reaction volume of 40 μl and mix (see REAGENT SETUP). Incubate in a thermal cycler for 15 min at 12 °C followed by 15 min at 25 °C.
▲ **CRITICAL STEP** Immediately proceed with the next step or put samples on ice after incubation. Extended incubation at room temperature (20–25 °C) may cause recessed instead of blunt ends.

SPRI purification ● TIMING 2 h

5| SPRI purification (solid-phase reversible immobilization) is used to clean reactions from salts, enzymes, nucleotides and unincorporated oligonucleotides. The DNA is reversibly bound to carboxyl-coated magnetic beads and released after washing away contaminants. Resuspend SPRI beads by vortexing or pipetting up and down. Add 72 μl of SPRI beads to each reaction. Vortex for several seconds to homogenize reactions and SPRI beads. Let stand at room temperature for 5 min. Collect the liquid at the bottom of the wells by briefly spinning to 2,000*g* in a plate centrifuge. Transfer the reaction plate onto a 96-well ring magnetic plate and let stand for 2 min to collect the beads on the walls of the wells. Carefully pipette off all supernatant without removing any beads.

6| While placed on the magnetic rack, wash the beads by adding 150 μl of freshly prepared 70% ethanol absolute. Incubate for 1 min and completely remove the ethanol without disturbing the beads. Repeat this step.

7| Air-dry the beads for 30–60 min at room temperature while leaving the lids open.

8| For elution add 22 μl EB and close the reaction wells with caps. Remove the plate from the magnetic rack and repeatedly vortex to resuspend the beads. Occasionally the beads may not fully resuspend, which does not negatively influence DNA recovery. Pull the liquid to the bottom of the wells by briefly spinning to 1,000*g* in a plate centrifuge. Transfer the plate back onto the 96-well ring magnetic plate and let stand for 1 min. Remove 20 μl of the eluate and transfer to a new 96-well reaction plate. Discard the remaining beads and liquid. Carryover of trace amounts of beads does not negatively affect subsequent reactions.
■ **PAUSE POINT** Purified samples can be stored at −20 °C for 1 d.

Ligation of barcoding adapters and SPRI purification ● TIMING 3 h

9| Add 20 μl of EB to one empty well on the reaction plate. This serves as a control for efficient removal of excess barcoding adapters during the subsequent purification steps.

10| Add 2 μl of barcoding adapter to each sample well. A different barcoding adapter must be added to each sample. For the positive and purification control reactions, arbitrary adapters may be chosen. Close the wells with caps and intensively vortex to mix the sample DNA and barcoding adapters. Collect the liquids at the bottom of the wells by briefly spinning to 1,000*g*.
▲ **CRITICAL STEP** Failure to mix samples and barcoding adapters properly before adding the ligation master mix may lead to increased chimera formation during ligation.

11| Prepare a ligation master mix for the required number of reactions. Because PEG is highly viscous, vortex the master mix before adding T4 ligase and mix gently thereafter.

Reagent	Volume (μl) per sample	Final concentration in reaction
Water	9	
T4 ligase buffer (10×)	4	1×
PEG-4000 (50%)	4	5%
T4 ligase (5 U μl^{-1})	1	0.125 U μl^{-1}

© 2008 Nature Publishing Group http://www.nature.com/natureprotocols

Illumina Exhibit 1014, Page 7 of 12

▲ CRITICAL STEP White precipitate may be present in the ligation buffer after thawing. Heat the buffer vial briefly to 37 °C and vortex until the precipitate has dissolved.

12| Add 18 µl of master mix to each reaction well to obtain a total reaction volume of 40 µl and mix. Visually verify that all reaction components are mixed well. Incubate for 1 h at 22 °C in a thermal cycler.

13| Perform SPRI purification exactly as described in Steps 5–8.
■ PAUSE POINT Purified samples can be stored at −20 °C for 1 d.

Adapter fill-in and SPRI purification ● TIMING 2.5 h
14| Prepare a master mix for the required number of samples.

Reagent	Volume (µl) per sample	Final concentration in reaction
Water	14.6	
ThermoPol buffer (10×)	4	1×
dNTPs (25 mM each)	0.4	250 µM each
Bst polymerase (8 U μl^{-1})	1	0.2 U μl^{-1}

15| Add 20 µl of master mix to each sample well to obtain a total reaction volume of 40 µl. Mix gently. Incubate at 37 °C for 20 min in a thermal cycler.

16| Perform SPRI purification exactly as described in Steps 5–8, but elute in TE instead of EB to stabilize the DNA for storage.
■ PAUSE POINT Barcoded samples can be stored at −20 °C for up to 1 week or at −80 °C for several months.

Verify tagging efficiency ● TIMING 1–2 h
17| Mix 7 µl of the tagged positive control PCR product and 4 µl of the untagged positive control PCR product with 2 µl loading dye each, and load them side-by-side with a size marker on a 2% agarose gel prepared in TBE buffer with 0.5 µg ml^{-1} ethidium bromide stain included in both the gel and the electrophoresis buffer. Apply a voltage of 5 V cm^{-1} for 40 min and visualize the bands under UV light. The positive control should produce a single sharp band, 40–50 bp longer than the untagged control (see **Fig. 3**). Remaining bands in the tagged positive control template that are the same size as the untagged positive control PCR product indicate inefficiency in either the blunt-end repair or the ligation step. Partly elevated bands indicate inefficiency or failure of the adapter fill-in step.
! CAUTION Handle ethidium bromide containing solutions only with gloves, protective clothing and proper precautions.
? TROUBLESHOOTING

Quantifying tagged samples ● TIMING 3–4 h
18| Measure the concentrations of the tagged samples using a PicoGreen quantification assay. Include both the positive control and the purification control in the measurements. The positive control should yield concentrations between 10 and 15 ng μl^{-1}, as the expected overall recovery from Step 2 through this step is between 40 and 60%. The negative control should not produce signals significantly above background.
▲ CRITICAL STEP Quantification accuracy affects the sequence representation of samples. Perform duplicate or triplicate measurements and be careful to obtain measurements within the linear range of your standard curve.
? TROUBLESHOOTING

Pooling tagged samples ● TIMING 1–2 h
19| Pool the tagged samples in molar ratios reflecting the proportion of sequence reads desired from each sample. For samples with similar mean fragment sizes, for example, samples that have been nebulized under the same conditions, pooling can be carried out using mass ratios that are then equivalent to molar ratios. If samples exhibit different mean fragment sizes, divide their mass concentration by the mean fragment size to obtain values proportional to their molar concentrations.

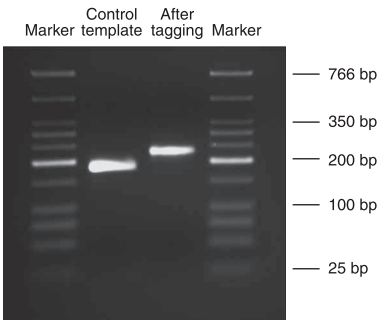

Figure 3 | Agarose gel electrophoresis to verify the efficiency of the tagging reactions. A 190 bp PCR product was used as the positive control template. A side-by-side comparison with the tagged PCR product shows the expected complete size shift of 44 bp, which corresponds to the size of two barcoding adapters ligated to both ends of the molecules.

© 2008 Nature Publishing Group http://www.nature.com/natureprotocols

Illumina Exhibit 1014, Page 8 of 12

▲ CRITICAL STEP Keep at least half of each tagged sample as backup, in case a second round of pooling and sequencing is required. Thus, half the volume of the sample with the lowest DNA concentration should be used for pooling, defining the volumes required for all other samples. If possible, the overall amount of DNA in the pool should exceed 10 ng, although as little as 1 ng of pooled material is usually sufficient to obtain a sequencing library for at least one complete 454 run.

20| If the volume of the pool exceeds 20 μl, concentrate the pool using the MinElute PCR purification kit according to the manufacturer's instructions. Elute in 20 μl EB.
■ PAUSE POINT The sample pool can be stored at −20 °C for up to 1 week.

Dephosphorylation and restriction digestion ● TIMING 2 h

21| For dephosphorylation, set up the following reaction in a single tube and mix gently.

Reagent	Volume (μl)	Final concentration in reaction
Water	14	
NEBuffer 3 (10×)	4	1×
Sample pool from Step 20	20	
Calf-intestine phosphatase (10 U μl^{-1})	2	0.5 U μl^{-1}

22| Incubate for 30 min at 37 °C in a thermal cycler. Purify the reaction using the MinElute PCR Purification kit. Elute in 20 μl EB.

23| For restriction digestion, set up the following reaction in a single tube and mix gently.

Reagent	Volume (μl)	Final concentration
Water	15.5	
Universal Buffer (10×)	4	1×
Eluate from Step 22	20	
*Srf*I (10 U μl^{-1})	0.5	0.125 U μl^{-1}

24| Incubate for 1 h at 37 °C in a thermal cycler. Purify the reaction using the MinElute PCR purification kit. Elute in 20 μl EB.
■ PAUSE POINT The eluate can be stored at −20 °C for up to 1 week.

Small fragment removal ● TIMING 1.5 h

25| The GS FLX sequencing platform produces read lengths of ∼250 bp. When sequencing shotgun libraries, fragments smaller than this should be removed to exploit the sequencing resources fully. Transfer the eluate from Step 24 into a PCR tube and add 20 μl EB for a total volume of 40 μl. Perform SPRI purification exactly as described in Steps 5–8, but reduce the volume of SPRI beads to 28 μl precisely. Elute in 20 μl EB.
▲ CRITICAL STEP Omit this step if molecules shorter than 250–300 bp should be sequenced, for example, short amplicons.
■ PAUSE POINT The eluate can be stored at −20 °C for up to 1 week.

454 Sequencing and data processing ● TIMING 2–3 d

26| Produce a 454 sequencing library using the GS DNA Library Preparation Kit. Use 20 μl of eluate from Step 24 or Step 25 and follow the manual starting at the adapter ligation step.

27| The sequencing libraries may be quantified by capillary gel electrophoresis or RiboGreen quantification as described in the GS DNA Library Preparation Kit. However, for better accuracy, we recommend using quantitative PCR as previously described[25]. Owing to insufficient sensitivity of the former methods, quantitative PCR is inevitable if little DNA (<500 ng) was available after pooling the tagged samples.
▲ CRITICAL STEP 454 Sequencing libraries rapidly degrade at room temperature. After quantification, immediately make aliquots and store them at −80 °C. Thaw single aliquots just before further processing with emulsion PCR and sequencing.
■ PAUSE POINT 454 Sequencing libraries can be stored at −80 °C for several months.

28| Proceed to emulsion PCR and sequencing according to the standard 454 routine (see ref. 3, and the GS FLX system manuals).

29| After sequencing, the 454 sequence reads must be sorted according to their sequence tags. Correctly tagged sequences start at the 5′ end with half an *Srf*I restriction site, GGGC, followed by a tag. Owing to frequent sequencing errors in and around homopolymer stretches, ambiguities should be allowed at the *Srf*I site. The following criteria give reliable results without

© 2008 Nature Publishing Group http://www.nature.com/natureprotocols

Illumina Exhibit 1014, Page 9 of 12

excessive rejection of reads: one to five Gs may be read, followed by one C, one or no additional G, and a perfectly matched tag. All remaining sequences are discarded. Software for this purpose is available online (http://bioinf.eva.mpg.de/pts/). It identifies and trims sequence tags and sorts the output into separate fasta or sff (standard flow format) files suitable for downstream analysis with the 454 assembly software. To assess data reliability, the false-assignment frequency should be estimated as described in **Box 1**.

? TROUBLESHOOTING

● TIMING

(For one plate with 94 samples and two controls)

Most of the steps include long incubation times, which can be used to set up and process additional plates in parallel.

Step 1: One to several days

Sample preparation (required time depends on the types of samples processed)

Steps 2–13: 1 d

Multichannel setup, blunt-end repair and reaction purification (Steps 2–8): 4–6 h

Ligation of barcoding adapters and reaction purification (Steps 9–13): 3 h

Steps 14–20: 1 or 2 d

Adapter fill-in and reaction purification (Steps 14–16): 2.5 h

Agarose gel electrophoresis (Step 17): 1–2 h

Quantification and pooling (Steps 18–20): 4–6 h

Steps 21–25: 3.5 h

Dephosphorylation and restriction digestion (Steps 21–24): 2 h

Small fragment removal (Step 25): 1.5 h

Steps 26–29: 3 d

Library preparation and quantification, emulsion PCR and 454 sequencing

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

TABLE 2 | Troubleshooting table.

Step	Problem	Possible reason	Solution
17	Positive control for the tagging reactions does not exhibit full size shift	Inactive enzymes, degraded buffers or dNTPs	Perform control reactions to check enzyme activity or repeat experiment with fresh reagents
		Improperly mixed reactions	Mix all reactions by flicking the reaction plate with a finger. Visually inspect the homogeneity of the reactions
		Excessive room temperature incubation at the blunt-end repair step	Place the reaction plate on ice after incubation and add SPRI beads while the reaction plate remains on ice
18	Positive control for the tagging reaction exhibits low concentration	Losses during SPRI purification	Take care to homogenize SPRI bead suspension and reaction properly by intensive vortexing
			Pipette carefully to avoid loss of beads during the washing steps
			While eluting the DNA, make sure all beads have come into contact with the elution buffer
	Negative control produces significant fluorescence signals	Incomplete removal of excess barcoding adapters	Perform an additional SPRI purification step. Take care to discard the supernatant completely while washing the SPRI beads. Excess barcoding adapters will not interfere with downstream reactions, but obscure the quantification signals and shift sequence representation among samples

© 2008 Nature Publishing Group http://www.nature.com/natureprotocols

Illumina Exhibit 1014, Page 10 of 12

TABLE 2 | Troubleshooting table (continued).

Step	Problem	Possible reason	Solution
29	A high proportion of sequences (>5%) does not carry a correct sequence tag	Dephosphorylation or restriction digestion failed	Perform control reactions to check enzyme activity or repeat experiment with fresh reagents
		Incorrect software usage	Manually inspect a subset of the sequences. Make sure you are using the right parameters as described in the manual
	High false-assignment rate	Barcoding adapters cross-contaminated	Order barcoding oligos from a different company and perform a small-scale control experiment
		Cross-contamination during reaction setup	Use filter tips
			Make sure to collect all liquids at the bottom of the wells before removing the caps
	Unequal sequence representation among samples	Inaccurate quantification of the samples before pooling	Repeat quantification in replicate measurements
			Make sure all measurements are in the linear range of your standard curve. Dilute one sample 1:1 with TE and check whether the quantity estimate decreases by half
			Re-pool the barcoded samples while correcting for the sequence representation observed and repeat sequencing
	Regions without coverage in shotgun assemblies	Average coverage too low	Sequence to higher average coverage
		Inaccurate pooling of long-range PCR products before nebulization	Sequence to higher coverage or requantify and pool the long-range PCR products
		*Srf*I restriction occurred within a sequence fragment	Use CpG methyltransferase to mask restriction sites within target regions

© 2008 Nature Publishing Group http://www.nature.com/natureprotocols

ANTICIPATED RESULTS

We have applied PTS to different types of samples, including PCR products and shotgun libraries derived from long-range PCR products and plasmids. When using 6- or 7-nt tags, we consistently find that 95–97% of the 454 sequences carry a correct sequence tag according to the criteria defined in Step 29 and can be assigned to their sample origin. **Figure 2** shows results we obtained from parallel sequencing 14 human mitochondrial genomes on a small GS FLX plate region. We recommend using a similar experimental design before large-scale adoption of PTS to test for possible cross-contamination of the barcoding adapters, estimate a false-assignment frequency and evaluate the sequence representation among samples. These parameters may vary slightly as they are affected by the synthesis quality of barcoding oligos, the sequencing platform (GS 20 or GS FLX), accuracy of quantification and handling.

Note: Supplementary information is available via the HTML version of this article.

ACKNOWLEDGMENTS We thank Christine Green for comments on the manuscript, Ellen Gunnarsdottir for help with providing data and Knut Finstermeier for help with the figures. This work was funded by the Max Planck Society.

Published online at http://www.natureprotocols.com
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* **74**, 5463–5467 (1977).
2. Hutchison, C.A. 3rd. DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res.* **35**, 6227–6237 (2007).
3. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
4. Bentley, D.R. Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.* **16**, 545–552 (2006).
5. Thomas, R.K. *et al.* Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picolitre reactor sequencing. *Nat. Med.* **12**, 852–855 (2006).

Illumina Exhibit 1014, Page 11 of 12

6. Poinar, H.N. *et al.* Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* **311**, 392–394 (2006).
7. Hofreuter, D. *et al.* Unique features of a highly pathogenic *Campylobacter jejuni* strain. *Infect. Immun.* **74**, 4694–4707 (2006).
8. Gowda, M. *et al.* Robust analysis of 5′-transcript ends (5′-RATE): a novel technique for transcriptome analysis and genome annotation. *Nucleic Acids Res.* **34**, e126 (2006).
9. Tawfik, D.S. & Griffiths, A.D. Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* **16**, 652–656 (1998).
10. Meyer, M., Stenzel, U., Myles, S., Prufer, K. & Hofreiter, M. Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Res.* **35**, e97 (2007).
11. Simcox, T.G. *et al.* SrfI, a new type-II restriction endonuclease that recognizes the octanucleotide sequence, (sequence: see text). *Gene* **109**, 121–123 (1991).
12. Pääbo, S. *et al.* Genetic analyses from ancient DNA. *Annu. Rev. Genet.* **38**, 645–679 (2004).
13. Willerslev, E. & Cooper, A. Ancient DNA. *Proc. Biol. Sci.* **272**, 3–16 (2005).
14. Roca, A.L., Georgiadis, N., Pecon-Slattery, J. & O'Brien, S.J. Genetic evidence for two species of elephant in Africa. *Science* **293**, 1473–1477 (2001).
15. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005).
16. Johnson, W.E. *et al.* The late Miocene radiation of modern Felidae: a genetic assessment. *Science* **311**, 73–77 (2006).
17. Nielsen, K.L., Hogh, A.L. & Emmersen, J. DeepSAGE—digital transcriptomics with high sensitivity, simple experimental protocol and multiplexing of samples. *Nucleic Acids Res.* **34**, e133 (2006).
18. Binladen, J. *et al.* The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PloS ONE* **2**, e197 (2007).
19. Hoffmann, C. *et al.* DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic Acids Res.* **35**, e91 (2007).
20. Parameswaran, P. *et al.* A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res.* **35**, e130 (2007).
21. Moore, M.J. *et al.* Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biol.* **6**, 17 (2006).
22. Wicker, T. *et al.* 454 sequencing put to the test using the complex genome of barley. *BMC Genomics* **7**, 275 (2006).
23. Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L. & Welch, D.M. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* **8**, R143 (2007).
24. DeAngelis, M.M., Wang, D.G. & Hawkins, T.L. Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Res.* **23**, 4742–4743 (1995).
25. Meyer, M. *et al.* From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Res*. 15 Dec. 2007 (Epub ahead of print).
26. Ahn, S.J., Costa, J. & Emanuel, J.R. PicoGreen quantitation of DNA: effective evaluation of samples pre- or post-PCR. *Nucleic Acids Res.* **24**, 2623–2625 (1996).
27. Singer, V.L., Jones, L.J., Yue, S.T. & Haugland, R.P. Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. *Anal. Biochem.* **249**, 228–238 (1997).

© 2008 Nature Publishing Group http://www.nature.com/natureprotocols

Illumina Exhibit 1014, Page 12 of 12