AMPLICON SEQUENCING using Illumina technology at the NSC

NORWEGIAN SEQUENCING CENTRE

www.sequencing.uio.no

This guideline is designed to help you initiate an amplicon sequencing project using our Illumina technology. Generally, amplicon-sequencing projects will be sequenced on our MiSeq system, but if greater yields are required, most samples will also be compatible with our NextSeq or HiSeq systems – please inquire to ous-seq@sequencing.uio.no

This guide contains the following sections:

1. **Introduction to Illumina sequencing technology.** It is important you fully understand the structure of Illumina adapter and primer sequences before attempting to design your own PCR primers for amplicon generation.
2. **Different amplicon sequencing strategies** provides an overview of methods to produce amplicon sequencing libraries.
3. **General considerations for producing amplicon libraries** provides essential information on producing and submitting amplicon libraries for sequencing to the NSC.
4. **16S/18S rDNA and ITS sequencing**
5. **Appendices 1 & 2:** single and dual index sequences used by Illumina.

DISCLAIMER

i. This document is intended as a guide only, and is not guaranteed error-free. There may exist alternative or superior improved methods not detailed herein, and users are advised to perform a literature review before initiating an amplicon-sequencing project. The inclusion of a method in this guide in no way implies that the NSC endorses the technique, nor that they have employed it for sequencing projects. *Users submitting their own libraries for sequencing assume the entire responsibility for any design errors, unless they have written confirmation from the NSC that the design has been conducted or approved by the NSC as part of an academic collaboration.*

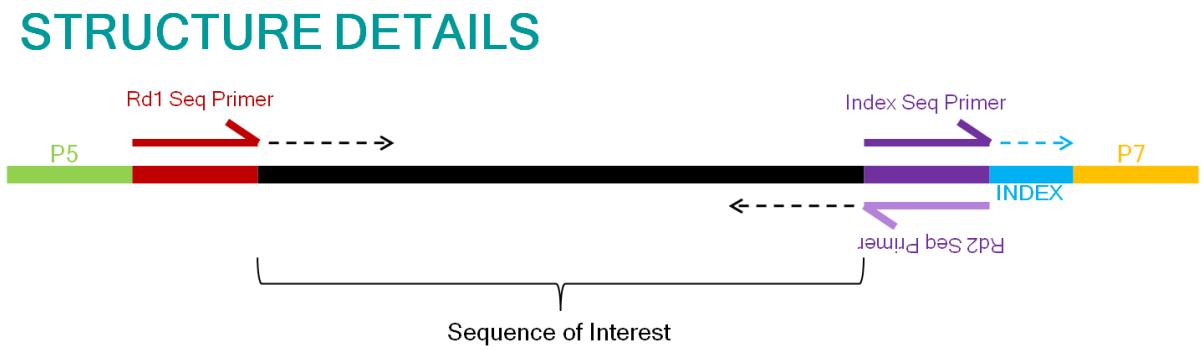
ii. Illumina does not support the use of custom sequencing primer sets. In the event of run failures during sequencing of custom libraries, Illumina may refuse to replace reagents expended. This is outwith the control of the NSC, and users will nonetheless be required to pay for reagents consumed.

1. INTRODUCTION TO ILLUMINA SEQUENCING TECHNOLOGY

Illumina library construction entails the addition of adapter sequences necessary for flow-cell binding and sequencing primer annealing to both ends of template molecules. Regardless of the sequencing primers and strategy employed (single read, paired end, and with or without index reads), libraries MUST contain the flow-cell binding sequences, called P5 and P7. The P5 and P7 sequence required for paired end sequencing is slightly longer than the minimum sequence which is sufficient for single-read flow-cells, so users are advised to use the longer paired-end compatible sequences – see details below on p10.

Structure of Illumina sequencing library constructs:

The figure below shows the structure of a typical Illumina library construct prepared for sequencing by the addition of flanking sequences. Flanking sequences can be added by ligation, by PCR with appropriately tailed primers, or by a combination of both ligation and PCR.

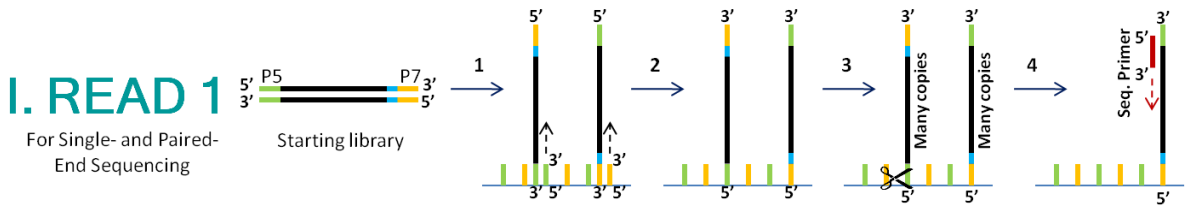


All figures courtesy of and used with permission of Harvard Medical School, Massachusetts General Hospital.

Sequencing of library constructs:

The figures below show the basic chemistry steps involved in Illumina sequencing:

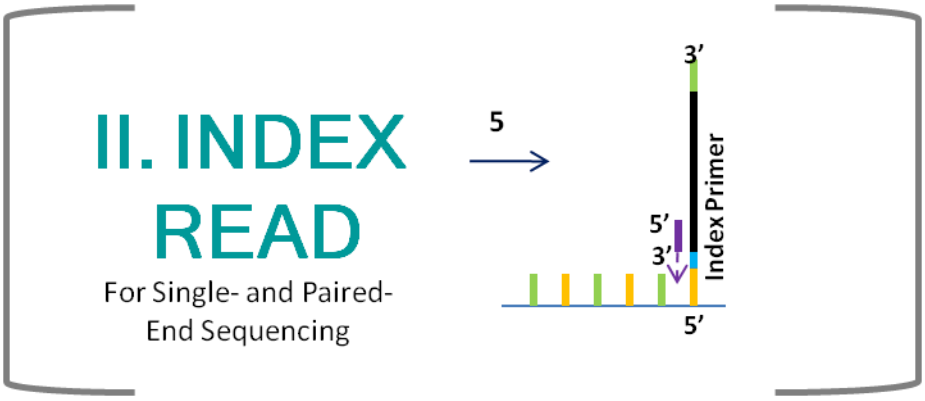
I. READ 1



1. The P5 and P7 regions of single-stranded library fragments anneal to their complimentary oligos on the flowcell surface. The flow cell oligos act as primers and a strand complimentary to the library fragment is synthesized.

2. The original strand is washed away, leaving behind complementary copies that are covalently bonded to the flowcell surface in a mixture of orientations.
3. Approximately 1,000 copies of each fragment are generated by bridge amplification, creating "clusters". For simplification, the diagram shows only one copy (out of 1,000) in each cluster, and only two clusters (out of 1-400 million, depending on machine used).
4. The P5 region is cleaved, resulting in clusters containing only fragments that are attached by the P7 region. This ensures that all copies are sequenced in the same direction. The sequencing primer anneals to the P5 end of the fragment, and begins the sequencing by synthesis process, which is repeated for a predefined number of cycles (50-300).

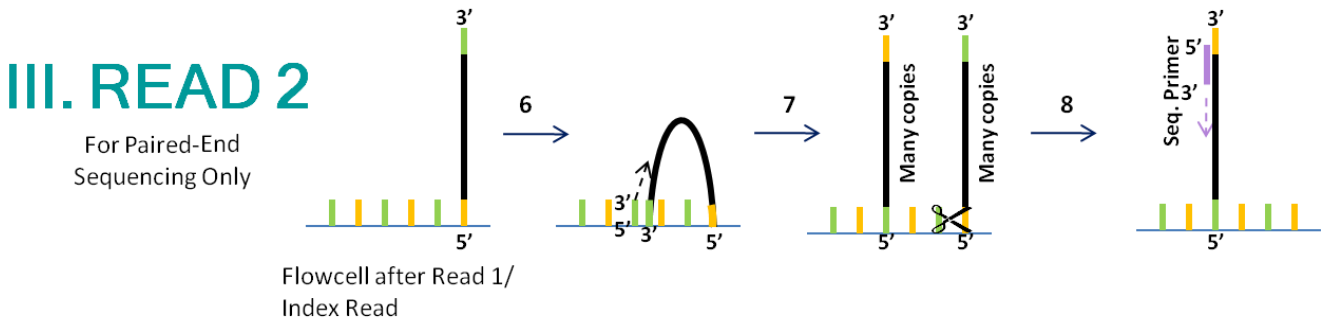
II. INDEX READS (optional)



5. Indexing (a.k.a. multiplexing, barcoding) allows different samples to be sequenced on the same lane of Illumina sequencing, with assignment of reads to the appropriate sample afterwards (called demultiplexing). When Read 1 is finished, everything from Read 1 is removed and an index primer is added, which anneals at the P7 end of the fragment and sequences the barcode.

6. If a dual indexing strategy is being followed to allow greater levels of sample multiplexing, a second index read is incorporated. The second index sequence (not shown in the figures here) is located at the P5 end of the construct, and the index read is primed from flow-cell lawn oligos, or a dedicated primer, depending on the machine used.

III. READ 2



7. Index read products are stripped from the template, which is then allowed to re-form

clusters by bridge amplification as in Read 1. This leaves behind fragment copies that are covalently bonded to the flowcell surface in a mixture of orientations.
8. This time, P7 is cut instead of P5, resulting in clusters containing only fragments that are attached by the P5 region. This again ensures that all copies are sequenced in the same direction (opposite from Read 1).
9. The read 2 sequencing primer anneals to the P7 region and sequences the other end of the template.

An animation of the above process can also be viewed here (as an example, as part of the Nextera workflow):
https://www.youtube.com/watch?v=womKfikWlxM&list=TLTdHTvoH8HscoJQpAHRqg71mBU6zDFPHQ

For more information, please see Illumina´s website: www.illumina.com

How does amplicon sequencing differ from regular DNA sequencing? (could also be tited "disadvantages of amplicon sequencing")

Regular genomic DNA sequencing usually uses a randomly fragmented library, thus sequencing reads will begin with a random selection of the bases A, G, C or T. In contrast, amplicon sequencing usually amplifies the same or similar DNA regions, with the result that most reads all begin with the same sequenced base. This is problematic for Illumina sequencing. In order to explain why, it is first necessary to explain some quirks of the sequencing process:

i. Cluster identification by Illumina´s imaging software requires an approximately equal balance of all 4 possible bases, through cycles 1-4.
ii. Estimation of error rates peculiar to sequencing-by-synthesis (phasing and prephasing) requires an approximately equal balance of all 4 possible bases through cycles 2-12.

It is worth noting that sequencing libraries derived from organisms with unusual GC:AT contents may cause problems due to the above requirements, regardless of whether amplicons or regular gDNA libraries are to be sequenced. Amplicons tend to be extreme cases, however.

The above requirements apply only to the READ1 and READ 2 sequencing reads, not the index reads. However, index reads have a requirement of their own:

i. Each sequencing cycle, irrespective of index read length, must contain at least one base from each of the following two pairs: (G/T) and (A/C). Ideally, there would be an approximately equal distribution of the bases.

If the above criteria cannot be satisfied (strategies for achieving this are covered below), the library with low diversity must be blended with a more complex library, up to a 1:1 ratio, to achieve a suitable level of diversity (lowering the possible yield of the target sequence to 50%). The library used for blending is usually from bacteriophage PhiX, which is purchased from Illumina and supplied by the NSC (alternatives are possible – see below). It is also often necessary to reduce the target yield to 80% normal levels, further reducing the possible yield from the target amplicon. For many

amplicon sequencing projects this is not an important consideration, as even 25% yields will saturate required sequencing levels, and the benefits of employing the simplest / cheapest library generation strategy outweigh those of lost yield.

Alternatives to PhiX. Any diverse library can be used to increase run diversity to acceptable levels, so you are encouraged to submit samples for regular gDNA library prep and sequencing simultaneously with your amplicon projects to maximise useful output. The NSC will pair users with compatible projects whenever possible, but this is rarely practical, so you are advised to seek partners in advance and make a combined submission**. A 1% blend of PhiX will be added to all runs for control purposes unless you explicitly ask for this to be omitted.**

Note on error rates

Error rates in your data are likely to be improved by using a higher proportion of PhiX blend, so if your application relies on accuracy (for example, detecting low-level SNPs in a population sample), you should request that we prioritise low error over yield (by blending more PhiX), and account for the lower yield in your calculations of number of runs required.

2. DIFFERENT AMPLICON SEQUENCING STRATEGIES

This section covers the generation of your own amplicon libraries by the following strategies:

		MiSeq yield possible
i.	AmpliSeq for Illumina – commercial primer panels	95%
ii.	Nextera XT / Nextera FLEX reagents from Illumina	95%
iii.	Simple amplicon production + standard Illumina library prep.	60 – 90%
iv.	Amplicon concatamerisation & fragmentation	95%
v.	Fusion primer library generation	60 – 90%
vi.	Custom fusion-primer sequencing libraries	60 – 90%

Note on pricing: Example pricing is given below where appropriate, but users should check with suppliers directly for up-to-date information. The NSC regrets it cannot purchase reagents on your behalf. Prices below cover library preparation only. Sequencing costs are in addition, for which you are encouraged to contact the NSC for an estimate on ous-seq@sequencing.uio.no

Note on read and amplicon lengths: The read lengths currently available on the MiSeq system are 50, 150, 250 and 300 bp, in both single-read and paired-end formats. Amplicons must therefore be shorter than 600 bp (2 x 300 bp paired end reads) in order to be sequenced in their entirety. However, longer amplicons can be fully sequenced if library generation using a fragmentation approach is employed (see strategies ii and iv).

i. Ampliseq for Illumina panels

Illumina sell a range of pre-made and custom primer panels with library generation reagents that will generate a library of sufficient diversity from which near 100% yields can be expected, using minimal PhiX blend. For more information see https://emea.illumina.com/products/by-brand/ampliseq.html

Price: Dependent on panel chosen and sample number. The NSC does not have the resources to assist users with designs.

ii. Nextera XT or Nextera Flex library kits from Illumina

Transposase mediated "tagmentation" based kits for Illumina library generation from up to 24 or 96 indexed samples. Nextera Flex reagents are less sensitive to input amount and require less amplification, but are more expensive than the older Nextera XT reagents. Both sets of reagents will produce a result similar to random fragmentation of the amplicon fragments, but with low to no coverage in the terminal 50 bases of each end. Amplicons longer than 500 bp can be sequenced by this strategy, but it is not recommended for amplicons shorter than 300 bp. Nextera libraries will be of sufficient diversity that near 100% yields can be expected, using minimal PhiX blend.

Price: e.g. Nextera XT kits of 24 samples ($895 ex MVA) or 96 samples ($3401 ex MVA). Also requires purchase of separate indexing reagents ($269 and $1061 ex MVA respectively). The NSC cannot purchase reagents on your behalf.

For more information, see:

https://emea.illumina.com/products/by-type/sequencing-kits/library-prep-kits/nextera-xt-dna.html?langsel=/no/

https://emea.illumina.com/products/by-type/sequencing-kits/library-prep-kits/nextera-dna-flex.html

iii. Simple amplicon production + standard Illumina library prep.

This is the simplest option for those with low numbers of amplicons to study. You send us purified PCR products, and we prepare a separate library from each, with indexing of up to 96 samples per lane of sequencing possible. Unless you are able to blend several amplicons such that the first 12 bases are an approximately equal balance of each of the 4 bases at each position, the low diversity requires that we blend as much as up to a 1:1 ratio of control library and reduce yield to 80 % as described above.

Price: 900-2400 NOK / sample (including MVA) depending on number of samples.

Another alternative (not currently offered as a service by the NSC) is to use a third-party reagent designed for converting PCR products into sequencing libraries, such as Qiagen's Qiaseq 1-step Amplicon Library Kit: https://www.qiagen.com/no/products/ngs/ngs-life-sciences/dna-amplicon-sequencing/

Variant 1 allowing up to 90% sequencing yield: If amplicons are designed with degenerate sequences at the 5´ends of both primers (12 N bases preceding specific priming sites), the resulting amplicons should have sufficient diversity to overcome the limitations of single amplicons. In practice, diversity is unlikely to equal that of randomly sheared DNA, so blending the amplicon library with control PhiX library will still be necessary, meaning it is impossible to achieve 100% possible yield.

Variant 2 allowing up to 90% sequencing yield: Primers with slightly different priming sites (varying by even just one 5´base at a time) can be designed to allow out-of-phase amplification of the same sequence, creating diversity. Although both ends of the amplicon contribute to the sequence diversity as seen by the Illumina sequencer, it can in practice be difficult to achieve an approximately equal distribution of all four nucleotides over the first 12 bases.

Variant 3 allowing up to 90% sequencing yield: Blending a number of different amplicons together (for example, multiple amplified loci from a single individual) prior to library generation can achieve the desired level of complexity.

iv. Amplicon concatamerisation & fragmentation

For studies with higher numbers of amplicons, or that require near-maximum yield (90%). You may ligate your amplicons into concatamers and submit the ligated concatamer(s) to us. We will then randomly fragment the concatamers by sonication prior to library generation, which increases sequence diversity in the first bases read. Up to 96 separate construct libraries can be prepared and run together in a single lane of sequencing by this method. Amplicons longer than 500 bp can be sequenced by this strategy.

1. **Price**: 900-2400 NOK / sample (including MVA) depending on number of samples.

v. Fusion primer library generation

Fusion primers allow both amplification of your target amplicons and simultaneous incorporation of the necessary sequences required for sequencing by Illumina technology, negating the need for subsequent library generation. This can be an economic solution, particularly if you will prepare libraries from the same primers many times. This section covers designs that utilise the *sequencing* primers employed by Illumina, generating libraries that are compatible with standard Illumina reagents. These libraries can be run together with samples from other users, and in the event of technical problems with sequencing hardware at the NSC, sequencing reagents will be replaced under guarantee. Users are responsible for design, and neither the NSC nor Illumina can accept responsibility for design errors.

All primer sequences below are given in 5´-3´orientation.

Sequences in lower case are required for Illumina flow cell binding and sequencing primer annealing sites. Sequencing primer annealing sites are underlined.

N = any nucleotide
X = gene specific primer sequence to be defined by user.
Z = P7 index sequence. Users are free to utilise index sequences of their choice, but an appendix of 48 recommended indexes can be found at the end of this document. Note that any index sequence you introduce here in a reverse primer must be the *reverse complement* of the sequence you intend to read as the index.

```
Forward primer:
aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctXXXXXXXXXXXXXXXXXXXX

Reverse primer:
caagcagaagacggcatacgagatZZZZZZgtgactggagttcagacgtgtgctcttccgatctXXXXXXXXXXXXXXXXXXX
```

For an example publication using this strategy, see Bartram et al, Appl. Environ. Microbiol. **June 2011** vol. 77 no. 11, pp **3846-3852.**

Variant 1 allowing higher sequencing yield (degenerate bases):

Use of the following alternative primers, incorporating 12 degenerate bases at the points where read 1 & 2 sequences will be generated, maximises diversity in these bases and allows the use of cluster numbers approaching 90% of maximum.

```
Forward primer:
aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctNNNNNNNNNNNNXXXXXXXXXXXXXX
XXXXXXX

Reverse primer:
caagcagaagacggcatacgagatZZZZZZgtgactggagttcagacgtgtgctcttccgatctNNNNNNNNNNNNXXXXXXX
XXXXXXXXXXXX
```


Variant 2 allowing higher yield (internal barcodes).

Similar to variant 1 above, but incorporating index sequences (also known as internal barcodes) in the first 12 bases of main read 1 and/or 2 (either as 12 base indexes, or shorter indexes in combination with Ns). Be aware that if using this approach, there must be sufficient diversity in the index sequences to allow almost equal proportions of each base at each of the 12 positions. Internal barcodes can be combined with standard third-read Illumina indexing to increase the number of samples that can be indexed in a single lane (see note on triple indexing below).

Variant 3: Two-step amplification.

The use of a 2-step PCR approach allowing the use of shorter primers for gene-specific amplification, each tagged with a fragment of the above sequence, is possible. A second amplification with another universal primer pair that extends the Illumina adapter sequences to full length is then applied. This use of shorter gene-specific primer pairs for initial amplification keeps primer costs to a minimum. This variant can in theory be combined with variant 1 or 2 above.

Variant 4: Two-step amplification utilising index-primer kits from Illumina (low-cost access up to 384 indexed primers)

When aiming to multiplex many samples, the requirement for many index primers can be prohibitively expensive. The solution can be a 2-step amplification strategy combining your own gene-specific primers with index-primer kits available from Illumina. Illumina describes this strategy in the following support bulletin – although written for 16S amplification, by replacing the 16S amplification primers with others of your choice, the method can be adapted for other sequences:

https://www.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf

This approach employs the Nextera™ tag sequences (note that these differ from the TruSeq-based sequences presented elsewhere in this guide) in combination with the Nextera XT™ Index Kit v2 index primers, available for purchase from Illumina for $1061 per box of 96 different indexes. Four different kits, A-D, are available, giving 384 possible index combinations.

Note: Achieving higher levels of multiplexing

Dual indexing to multiplex greater sample numbers whilst keeping primer costs to a minimum:

Indexes can also be incorporated at the P5 end of the sequencing library. P5 indexing should not be used in isolation, always in combination with P7 indexes. Dual indexing with a combination of P5 and P7 indexes allows a greater number of samples to be multiplexed with only a few index primers – e.g. combining eight P5 indexes with twelve P7 indexes gives 96 possible combinations. Index combinations must be chosen carefully to ensure compatibility – please consult published studies. A list of indexes commonly employed for dual indexing by Illumina is provided in Appendix 2. If employing dual indexing, the P5 index sequence should be inserted as follows (ZZZZZZZZ = index):


```
Forward primer:
aatgatacggcgaccaccgagatctacacZZZZZZZZacactctttccctacacgacgctcttccgatctXXXXXXXXXXXXX
```

Triple indexing:

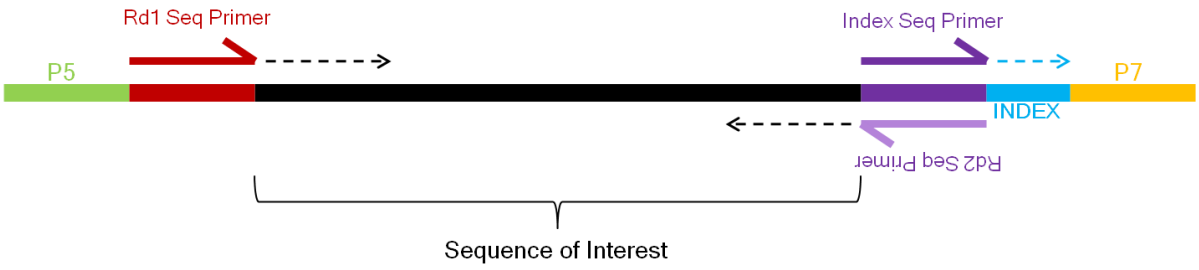
Greater levels of indexing can be achieved by combining index reads with “barcode” indexes incorporated in the main sequencing reads. Please see for example https://www.ncbi.nlm.nih.gov/pubmed/28683838

vi. Custom fusion primer sequencing libraries

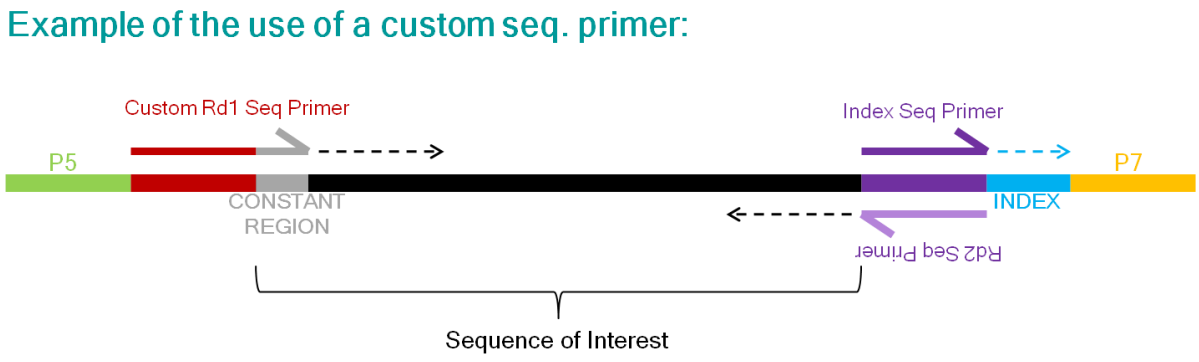
Illumina does not support the use of custom sequencing primers, but there are circumstances when the benefits of their use outweigh the potential risks of lost sequencing runs. Designed correctly, custom sequencing primers can allow the user to overcome the constraints of lower sequence yield caused by sequencing reads beginning with a constant gene-specific priming region, or allow the use of shorter primers. This requires design of alternative sequencing primers from those employed by Illumina, but an absolute requirement is that the P5 and P7 regions required for flow-cell binding be maintained.

Compare the regular Illumina library construct with the example custom sequencing (read 1 only in this example) product below:

Regular library construct:



Custom library construct (custom read 1 sequencing primer):



More complex variants involving the use of custom read1, index read and read 2 sequencing primers are also possible. For an example, please see the following publication: Caporaso et al (2012), ISME J., 6(8):1621-4. In order to allow maximum output from the MiSeq, sequences read from the read 1 and read 2 sequencing primers must have approximately random sequence (equal occurrence of the 4 bases) in the 12 first sequenced base positions. Should a third read indexing strategy be pursued, this must also conform to the recommendations for index reads (see instructions on p4).

Designing custom amplicon primers

An absolute requirement is that the 5´termini of amplicons contain the P5 and P7 sequences below necessary for flow cell binding. These sequences are only a piece of the full adapter construct required for Illumina sequencing, so you will need to add sequencing primer binding sites and site-specific priming sites as a minimum, and also index read sequences and primer binding site if desired.

Minimum P5 / P7 sequences required for single-read ONLY flow cells

```
P5: 5' AATGATACGGCGACCACCGA 3'
P7: 5' CAAGCAGAAGACGGCATACGA 3'
```

P5 / P7 sequences required for single-read & paired-end compatible flow cells (RECOMMENDED)

```
P5: 5' AATGATACGGCGACCACCGAGATCTACAC 3'
P7: 5' CAAGCAGAAGACGGCATACGAGAT 3'
```


Custom sequencing primers should also:

- For read 1, anneal to the P5 end of the library.
- For index read (if using optional 3rd read indexing), anneal to the P7 end of the library in the same orientation as the read 1 primer (See figure above).
- For read 2, anneal to the P7 end of the library (in reverse complement to other primers –see figure above).
- Should span any constant / low diversity sequence regions if you wish to obtain maximum sequence output.
- Be free of secondary structures (hairpins, loops etc)
- Have Tm and GC % similar to Illumina´s sequencing primers. Since Tm calculation methods differ, the sequence of the original sequencing primers is provided below for reference:

```
Read 1 sequencing primer:     ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Index sequencing primer:      GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
Read 2 sequencing primer:     GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
```

Custom sequencing primers should be provided to us as follows:

- HPLC purified
- Resuspended in EB (10 mM Tris) at 100 µM (100 pmol/µl)
- Minimum 15 µl must be provided.

Custom sequencing primer validation

Despite careful design, there is no guarantee that custom sequencing primers will perform successfully during Illumina sequencing reactions. The best test you can perform prior to submission is to perform a test PCR reaction, using your amplicon as template and your sequencing primers in appropriate combination with additional primers representing the P5 and p7 sequences. Truncated PCR product will be generated if your sequencing primer and P5 /P7 anneals effectively. However, this unfortunately does not guarantee equivalent performance in the MiSeq itself, which uses proprietary reaction mixes. It should also be noted that primer combinations which may have functioned well in the older Illumina GAIIx or HiSeq systems will not necessarily function in the MiSeq, as different chemistries and temperatures are employed between these systems.

You may also apply to Illumina for full details of the primer and adapter sequences they employ, which may help you design your own constructs. Please contact **customerservice@illumina.com**

Compatibility & design constraints for custom amplicon sequencing at the NSC

Finally, please note that we usually only accept custom sequencing primer projects for the MiSeq or NextSeq platforms. These are generally not compatible for combination with other user´s projects on our HiSeq machines. However, if you expect to have an extremely large project that could occupy an entire flow cell of the HiSeq, we will consider this. Contact us on ous-seq@sequencing.uio.no

NSC employees do not have the capacity to aid you in design, although we may consider this in exceptional cases if performed as an academic collaboration.

3. GENERAL CONSIDERATIONS FOR PRODUCING AMPLICON LIBRARIES

1. A high-fidelity polymerase should be used in the amplicon generation step.

2. You must purify PCR products prior to submission for sequencing. PCR products must be free of primer dimer products. If you have trouble removing these with cleanup columns, the use of AMPure XP beads (Beckman Coulter) may help. For large dimer products, reducing the ratio of AMPure beads to DNA may be necessary (for example, reducing the ratio of beads to DNA from 1.8:1 to 1:1).

3. When sequencing mixtures of multiple amplicons, care must be taken in quantification and pooling of amplicons. Amplicons must be cleaned up following PCR, and free from primer

dimers, prior to quantification. Equimolar mixtures of amplicons will give best results if equal coverage is desired. For quantification, we recommend the Qubit system (Invitrogen) or similar fluorometric system.

4. For minimum concentrations and volumes required for sequencing, please see our submission guidelines:
   http://www.sequencing.uio.no/forms/guidelines-submission-form-(illumina).pdf

How to submit samples: When you have prepared your amplicons for sequencing by any of the methods described above, please complete our regular submission form available from the link below. It is essential you also read the accompanying guidelines.

http://www.sequencing.uio.no/forms/samplesubmissionforms.html
http://www.sequencing.uio.no/forms/guidelines-submission-form-(illumina).pdf

4. 16S, 18S and ITS sequencing

16S analysis is amongst the most challenging applications of Illumina sequencing, as it requires sequencing of amplicons containing low-diversity "constant" regions, whilst determining accurate base calls over "variable" regions. All 16S projects therefore must be run at reduced MiSeq loading (4/5^{ths} max.) typically with a 20% PhiX (or equivalent diverse library) blend. We currently recommend that you follow the Fadrosh et al. protocol (see link below), which allows the use of only 10% PhiX blend and in our experience delivers higher read quality than other procedures:

http://www.microbiomejournal.com/content/2/1/6

Illumina has also produced the following application note detailing an alternative procedure. Libraries generated using this method will require a 20% PhiX blend. Please note that the NSC does not offer analysis using the MiSeq Reporter Software as part of our service:

https://www.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf

If you will only be sequencing a few samples, the investment in primers required may be prohibitively expensive, in which case a commercial kit may be more appropriate. See for example:

http://www.biooscientific.com/Illumina-Metagenomics-Library-Prep-kits (20% PhiX blend required)
https://www.qiagen.com/no/ (16S and fungal ITS kits available, only 10% PhiX blend required)
https://swiftbiosci.com/products/new-swift-amplicon-16sits-panel/ (16S and ITS kits available, only 10% PhiX blend required)

Ultra-throughput 16S sequencing, appropriate for projects with thousands of samples, can also be performed on the HiSeq 2500 or NovaSeq systems with 250 bp paired end reads. Examples of demonstrated protocols can be found here:

https://www.ncbi.nlm.nih.gov/pubmed/28683838
https://msystems.asm.org/content/4/1/e00029-19

5. APPENDIX 1 – P7 single indexes

Suggested 6 bp index sequences.

The following index sequences are commonly used in single-index adapters. Note that if you incorporate these or any other indexes of your choice into a reverse fusion primer for amplicon generation, the ***reverse complement of the sequences detailed here*** must be entered into the primer. The sequences given below are what will be read and output by the Illumina sequencer, and are what we require you to provide to us in detailing the indexes used.

ATCACG	index 1	AGTCAA	index 13	ACTGAT	index 25	CGGAAT	index 37
CGATGT	index 2	AGTTCC	index 14	ATGAGC	index 26	CTAGCT	index 38
TTAGGC	index 3	ATGTCA	index 15	ATTCCT	index 27	CTATAC	index 39
TGACCA	index 4	CCGTCC	index 16	CAAAAG	index 28	CTCAGA	index 40
ACAGTG	index 5	GTAGAG	index 17	CAACTA	index 29	GACGAC	index 41
GCCAAT	index 6	GTCCGC	index 18	CACCGG	index 30	TAATCG	index 42
CAGATC	index 7	GTGAAA	index 19	CACGAT	index 31	TACAGC	index 43
ACTTGA	index 8	GTGGCC	index 20	CACTCA	index 32	TATAAT	index 44
GATCAG	index 9	GTTTCG	index 21	CAGGCG	index 33	TCATTC	index 45
TAGCTT	index 10	CGTACG	index 22	CATGGC	index 34	TCCCGA	index 46
GGCTAC	index 11	GAGTGG	index 23	CATTTT	index 35	TCGAAG	index 47
CTTGTA	index 12	GGTAGC	index 24	CCAACA	index 36	TCGGCA	index 48

Note that when selecting index combinations, each cycle, irrespective of index read length, must contain at least one base from each of the following two pairs: (G/T) and (A/C). For example, if using only two indexes, index 6 and 12 are compatible.

If using other indexes, it is your responsibility to ensure that all sequences differ by at least 3 nucleotides (allows demultiplexing permitting 1 mismatch), or inform us to otherwise demultiplex allowing no mismatches.

APPENDIX 2 – P7/P5 dual indexes

Suggested 8 bp sequences for dual indexing

Index read 1 (P7) sequences. Note that if you incorporate these or any other indexes of your choice into a reverse fusion primer for amplicon generation, the ***reverse complement of the sequences detailed here*** must be entered into the primer.

ATTACTCG	D701
TCCGGAGA	D702
CGCTCATT	D703
GAGATTCC	D704
ATTCAGAA	D705
GAATTCGT	D706
CTGAAGCT	D707
TAATGCGC	D708
CGGCTATG	D709
TCCGCGAA	D710
TCTCGCGC	D711
AGCGATAG	D712

Index read 2 (P5) sequences (should be entered as shown here into the forward primer sequence).

TATAGCCT	D501
ATAGAGGC	D502
CCTATCCT	D503
GGCTCTGA	D504
AGGCGAAG	D505
TAATCTTA	D506
CAGGACGT	D507
GTACTGAC	D508