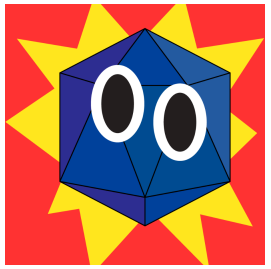
Apr 28, 2020

PrimalSeq: Generation of tiled virus amplicons for MiSeq sequencing



In 1 collection

DOI
dx.doi.org/10.17504/protocols.io.bez7jf9n

Nate Matteson[1], Nathan D Grubaugh[2], Karthik Gangavarapu[1], Josh Quick[3], Nick Loman[3], Kristian Andersen[1]

[1]Scripps Research, Department of Immunology and Microbiology;
[2]Department of Epidemiology of Microbial Diseases, Yale School of Public Health;
[3]University of Birmingham, Institute of Microbiology and Infection

Nate Matteson

OPEN ACCESS

DOI: dx.doi.org/10.17504/protocols.io.bez7jf9n

External link: https://docs.google.com/document/d/1zdUu9DPLT38Ize-SkQL6UOKKhcwsFciyh8p3-HaOWEg/edit?usp=sharing

Protocol Citation: Nate Matteson, Nathan D Grubaugh, Karthik Gangavarapu, Josh Quick, Nick Loman, Kristian Andersen 2020. PrimalSeq: Generation of tiled virus amplicons for MiSeq sequencing. **protocols.io https://dx.doi.org/10.17504/protocols.io.bez7jf9n**

Manuscript citation:
Grubaugh, ND. et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biology 20,8 (2019) https://genomebiology.biomedcentral.com/articles/10.1186/s13059-018-1618-7

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working
We use this protocol readily in our lab. It is successful for a wide range of virus samples.

Created: April 14, 2020

Last Modified: April 28, 2020

Protocol Integer ID: 35615

Abstract

Generated in collaboration by the Loman, Andersen, and Grubaugh labs.

For general use of the protocol and primer design, please cite:
Quick, J. et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature Protocols* 12 (6), 1261-1276 (2017)https://www.nature.com/nprot/journal/v12/n6/abs/nprot.2017.066.html

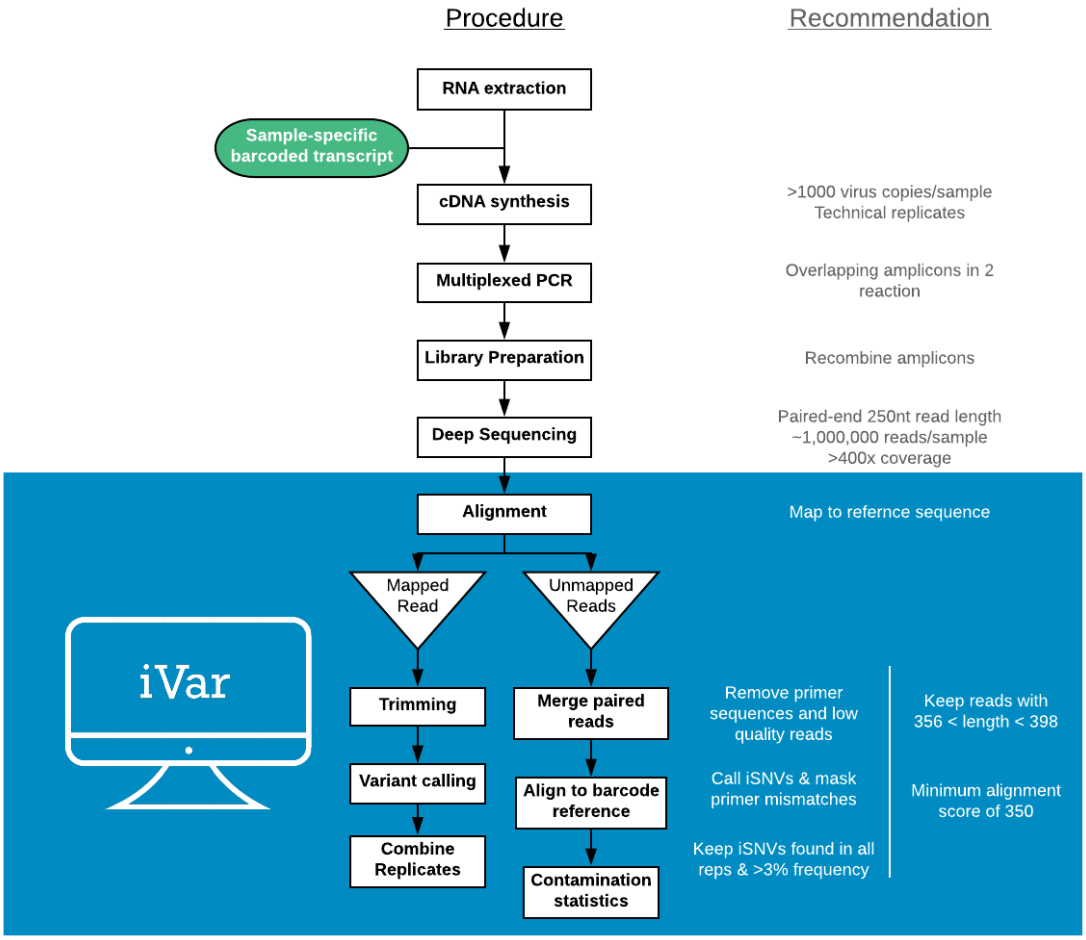
For measuring intrahost virus genetic diversity and calling variants using iVar, please cite:
Grubaugh, ND. et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biology* 20,8 (2019)
https://genomebiology.biomedcentral.com/articles/10.1186/s13059-018-1618-7

The general approach to this protocol is to amplify the virus genome in small (~400 bp) overlapping fragments using two highly multiplexed PCR reactions (where the overlapping segments are in separate reactions). The amplicons are combined after PCR and are the correct size for library preparation and paired-end 250 nt sequencing using the Illumina MiSeq.

Version 4 notes: This protocol has been updated to include considerations for measuring intra-batch contamination through the introduction of sample-specific barcoded spike-ins.

All of the primers are now listed in a separate spreadsheet. We currently have 400 bp amplicon schemes for Zika virus, West Nile virus (North America lineage I genotype), Usutu virus, and chikungunya virus (ECSA genotype). More will be made available soon. You can build your own primer sets by using **Primal Scheme**.

Overview of tiled virus amplicon sequencing protocol


Procedure
Recommendation
RNA extraction
Sample-specific barcoded transcript
cDNA synthesis
>1000 virus copies/sample
Technical replicates
Multiplexed PCR
Overlapping amplicons in 2 reaction
Library Preparation
Recombine amplicons
Deep Sequencing
Paired-end 250nt read length
~1,000,000 reads/sample
>400x coverage
Alignment
Map to refernce sequence
Mapped Read
Unmapped Reads
iVar
Trimming
Variant calling
Combine Replicates
Merge paired reads
Align to barcode reference
Contamination statistics
Remove primer sequences and low quality reads
Call iSNVs & mask primer mismatches
Keep iSNVs found in all reps & >3% frequency
Keep reads with 356 < length < 398
Minimum alignment score of 350

Guidelines

Considerations for measuring intrahost genetic diversity using this amplicon-based protocol:

1. Requires at least 1000 virus RNA copies going into cDNA synthesis. More is better. Try to normalize virus RNA copies between samples to make comparisons easier.
2. Process each RNA sample twice through the protocol to sequence as technical replicates. By calling variants only present in both replicates, it reduces the number of false positives (mainly from sequencing errors) and increases the accuracy of variant frequency measurements.
3. Obtain at least 400x nt coverage of each nucleotide position. Because of different amplicon efficiencies, this typically means that ~1M 250 nt paired-end reads are needed. Amplification of high input virus concentrations (>10,000 virus RNA copies) are more even and require fewer total reads.
4. During our validation process, the lowest intrahost variant frequency that we could accurately and consistently measure was 3%. Measuring lower than this requires additional input copies, coverage depth, and validation.
5. Beware of intrahost virus variants that exist within primer binding sites as they can decrease the amplification efficiency of that particular virus haplotype. Because the primer sites are trimmed and are covered by an overlapping amplicon, the variants within the primer sites can be accurately measured. All variants within the amplicon with a primer mismatch, however, can be significantly altered. This is the major limitation with any PCR protocol for virus population diversity analysis.
6. Use our data pipeline, iVar (intrahost variant analysis from replicates) to process and analyze the data. It will align to the reference (or call a consensus), trim primers, call variants, compare variants between replicates, and flag variants within primer sites.

Considerations for estimating intra-batch contamination using this amplicon-based protocol:

1. Requires a suitable amount of barcode reads. We aim to get 1% barcode reads for each sample which, given our recommendation of ~1M reads per sample, amounts to ~10,000 barcode reads. Less than this and small amounts of spillover have exaggerated effects. For RNA extracted from mosquito pools, we find an addition of 10 fg of spike-in to be suitable for most samples.
2. Limit of detection at 1% barcode reads was found to be ~0.05% contaminating reads if spiked-in barcode transcripts have completely different barcode regions and ~0.1% if barcode transcripts share one barcode region.

Materials

MATERIALS

- Q5 High-Fidelity 2X Master Mix - 100 rxns **New England Biolabs** **Catalog #**M0492S
- Agencourt AMPure XP **Beckman Coulter** **Catalog #**A63880
- KAPA Hyper Prep Kits (24 rxns with amplification) **Kapa Biosystems** **Catalog #**07962347001
- Agilent High Sensitivity DNA Kit **Agilent Technologies** **Catalog #**5067-4626
- Qubit 1X dsDNA High Sensitivity Assay Kit **Thermo Fisher Scientific** **Catalog #**Q33230
- KAPA Library Quantification Kit for Illumina® Platforms **Kapa Biosystems** **Catalog #**KK4835
- SuperScript™ IV VILO™ Master Mix **Thermo Fisher** **Catalog #**11756500
- Custom primers
- BIOO Scientific NEXTflex Dual-indexed DNA Barcodes

Generation of barcode spike-ins

1 Barcode spike-ins are generated with a step-out PCR using a common insert and variable forward and reverse primers. The insert we use and structures of the variable primers is show below. X's refer to the barcode region. We generated 8 forward barcodes, and 12 reverse barcodes which are 20bp long and have an edit distance of at least 10 nucleotides from all other generated barcodes. By mixing and matching primers a total of 96 unique barcode transcripts can be generated.

```
> Zea mays alcohol dehydrogenase 1 (adh1), mRNA (XM_008650471.2)
AGGGTCTCGGAGTGGATTGATTTGGGATTCTGTTCGAAGATTTGCGGAGGGGGGCAATGGCGACCG
CGGGGAAGGTGATCAAGTGCAAAGCTGCGGTGGCATGGGAGTCCATGCAAGCCACTGTCGATCGAG
GAAGTGGAGGTAGCGCCTTCGCAGGCCATGGAGGTGCGCGTCAAGATCCTCTTCACCTCGCTCTGC
CACACCGACGTCTACTTCTGGGAGGCCAAGGGGCAGACTCCCGTGTTCCCTCGGATCTTTGGCCAC
GAGGCTGGAGGTATCATAGAGAGTGTTGGAGAGGGTGTGACTGACGTAGCTC

> Barcode transcript step-out forward primer (structure)
AATGTCGCAGGCACTTGTCCxxxxxxxxxxxxxxxxxxxxAGGGTCTCGGAGTGGATTGA

> Barcode transcript step-out reverse primer (structure)
GGTCAGAGCTGTCTCCTGCTxxxxxxxxxxxxxxxxxxxxGAGCTACGTCAGTCACACCC
```

2 Prepare a PCR reaction for each combination of forward and reverse primers. A master mix can be created by combining all components except the forward and reverse primers.

	Component	Volume in 20 µL reaction	Final Concentration
	Q5 Reaction Buffer (5x)	5 µL	1x
	dNTPs (10 mM)	0.5 µL	200 uM
	Forward Primer (10 µM)	1 µL	400 nM
	Reverse Primer (10 µM)	1 µL	400 nM
	Q5 Polymerase	0.25 µL	0.02 U/ul
	Water	16.25 µL	-
	Template DNA (adh1; 0.1 ng/µL)	1 µL	<1000 ng

3 Run the following cycles on a thermocycler:

	Cycles	Temperature	Time
	1	98°C	30 seconds
	10	98°C	10 seconds
		68°C	10 seconds
		72°C	20 minutes
	1	72°C	2 minutes
	1	4°C	∞

4 Proceed immediately to cleanup

Post PCR cleanup

5 Allow Mag-Bind TotalPure NGS beads to equilibrate to room temperature, vortex until homogenous.

6 Bring PCR product volume up to 25 µL with water (if not at volume already).

7 Add 50 µL of beads to 25 µL of PCR product, mix well, and incubate at room temperature for 10 minutes.

8 Place tubes on a magnetic stand and incubate until solution appears clear.

9 Discard supernatant without disturbing the beads.

10 While tubes are on the magnet, add 200 µL of 80% EtOH, incubate for 30 seconds, and discard the EtOH wash.

11 Repeat previous 80% EtOH wash and remove as much EtOH as possible.

12 Leave tubes on magnet and air dry for 5 minutes.

13 Remove tubes from magnet and add 20 µL of nuclease-free water. Mix well by pipetting.

14 Place tubes on magnet stand. When solution appears clear, remove supernatant without disturbing the beads and place into new tubes.

15 Quantify the DNA concentration using the Qubit High Sensitivity DNA kit (or equivalent) from 1 µL of each product. Expected range = 10-100 ng/µL DNA. Sequencing from lower concentrations may still work.

16 *Note: If your lab has a KingFisher, you can download our automated protocols here*: https://github.com/grubaughlab/Kingfisher_protocols (use 'purification.bdz' for this step)

Preparation of cDNA

17 Isolate viral RNA using Omega Viral DNA/RNA kit, Trizol, or equivalent.

18 Many different cDNA synthesis kits can be used, but choose something that is relatively high-fidelity. The current protocols uses SuperScript IV VILO Master Mix because the enzyme has low error rates and the protocol is fast and easy.

19 Dilute a working stock of Barcoded Spike-ins 1:100,000 to obtain a concentration of 38 fM. Select unique spike-in for each sample. Try not to repeat spike-ins from recent runs. *Note: Be careful to not cross-contaminate the spike-ins by centrifuging all liquid from the caps and only opening one index at a time*

20

	Component	Volume in 20 µL reaction
	SSIV VILO Master Mix	4 µL

	Nuclease-free water	5-14 µL
	Barcoded Spike-in (38 fM)	1 µL
	Virus RNA	1-10 µL

21 Run the following cycles on a thermocycler:

	Temperature	Time
	25°C	10 minutes
	50°C	10 minutes
	85°C	5 minutes
	4°C	∞

22 Store samples at 4°C (for use same day) or -20°C (for use within a week) until ready for PCR.

PCR generation of tiled amplicons

23 Validated primer schemes can be found **here**. Prepare two primer pools by mixing equal volumes of each 10 µM primer. Primers indicated by "*" should be pooled at a concentration of 50 µM and primers indicated by "**" should be pooled at a concentration of 100 µM to help normalize sequencing coverage. The sequences for the primer which amplify the barcoded transcript are shown below:

```
> Contamination Primer Forward
AATGTCGCAGGCACTTGTCC

> Contamination Primer Reverse
GGTCAGAGCTGTCTCCTGCT
```

Note: Concentration for contamination primers should be identical to concentration of individual primer pairs in pool. Concentration listed below is for our West Nile virus protocol but will vary based on other schemes.

24 Prepare two PCR reactions for each sample (one for each primer pool):

	Component	Volume in 25 µl reaction
	Q5 2x Master Mix	12.5 µl
	Primer pool (#1 or #2)	1 µl
	Contamination Primer Forward (0.26 µM)	1 µl
	Contamination Primer Reverse (0.26 µM)	1 µl
	Nuclease-free water	8.5 µl
	cDNA	1 µl

25 Run the following cycles on a thermocycler

	Cycles	Temperature	Time
	1	98°C	30 seconds
	35	95°C	15 seconds
		65°C	5 minut

			es
	1	4°C	∞

26 Run 5 µl of each product on a 1% agarose gel. Each should produce a visible 400 bp band.

Post PCR cleanup

27 Allow Mag-Bind TotalPure NGS beads to equilibrate to room temperature, vortex until homogenous.

28 Bring PCR product volume up to 25 µL with water (if not at volume already).

29 Add 45 µL of beads to 25 µL of PCR product, mix well, and incubate at room temperature for 10 minutes.

30 Place tubes on a magnetic stand and incubate until solution appears clear.

31 Discard supernatant without disturbing the beads.

32 While tubes are on the magnet, add 200 µL of 80% EtOH, incubate for 30 seconds, and discard the EtOH wash.

33 Repeat previous 80% EtOH wash and remove as much EtOH as possible.

34 Leave tubes on magnet and air dry for 5 minutes.

35 Remove tubes from magnet and add 20 µL of nuclease-free water. Mix well by pipetting.

36 Place tubes on magnet stand. When solution appears clear, remove supernatant without disturbing the beads and place into new tubes.

37 Quantify the DNA concentration using the Qubit High Sensitivity DNA kit (or equivalent) from 1 µL of each product. Expected range = 10-100 ng/µL DNA. Sequencing from lower concentrations may still work.

38 *Note: If your lab has a KingFisher, you can download our automated protocols here*: https://github.com/grubaughlab/Kingfisher_protocols (use 'purification.bdz' for this step)

End-repair and A-tailing

39 Combine 25-50 ng of PCR-amplified DNA from primer pool 1 and 2 together for a total of 50-100 ng in 12.5 µL (equal concentrations of each amplicon pool). QS to a total volume of 12.5 µL using nuclease-free water.

Alternatively: proceed using 50-100 ng of primer pool product separately for library preparation. This allows for additional monitoring of cross-contamination. Data can be merged computationally post sequencing.

40 Combine the following components from the Kapa Hyper prep kit for end repair:

	Component	Volume in 15 µl reaction
	End Repair & A-tailing buffer	1.75 µl
	End Repair & A-tailing enzyme mix	0.75 µl
	PCR-amplified DNA (50 ng)	12.5 µl

41 Run the following cycles on a thermocycler:

	Temperature	Time

	20°C	30 minutes
	65°C	30 minutes
	4°C	∞

Adapter ligation

42 Dilute a working stock of NEXTflex Dual-Indexed DNA Barcodes 1:100 to obtain a concentration of 250 nM. Select unique barcodes for each sample. Try not to repeat barcodes from recent runs.

Note: Be careful to not cross-contaminate the adaptors by centrifuging all liquid from the caps and only opening one index at a time.

43 Combine the following components:

	Component	**Volume in 27.5 µL reaction**
	Ligation buffer	7.5 µL
	DNA ligase	2.5 µL
	NEXTflex DNA Barcodes (250nM)	2.5 µL
	End repair reaction product	15 µL

44 Incubate at 20°C for 15 minutes

45 Proceed immediately to cleanup

Post ligation cleanup

46 Allow Mag-Bind TotalPure NGS beads to equilibrate to room temperature, vortex until homogenous.

47 Add 22 µL of beads to 27.5 µL of ligation product, mix well, and incubate at room temperature for 10 minutes.

48 Place tubes on a magnetic stand and incubate until solution appears clear.

49 Discard supernatant without disturbing the beads.

50 While tubes are on the magnet, add 200 µL of 80% EtOH, incubate for 30 seconds, and discard the EtOH wash.

51 Repeat previous 80% EtOH wash and remove as much EtOH as possible.

52 Leave tubes on magnet and air dry for 5 minutes.

53 Remove tubes from magnet and add 20 µL of nuclease-free water. Mix well by pipetting.

54 Place tubes on magnet stand. When solution appears clear, remove supernatant without disturbing the beads and place into new tubes - 15 µL will go into library amplification.

55 *Note: If your lab has a KingFisher, you can download our automated protocols here*: https://github.com/grubaughlab/Kingfisher_protocols (use 'purification.bdz' for this step)

Library amplification

56 Combine the following components:

	Com pone	Volu me in

	nt	34 µL reaction
	2X KAPA HiFi HotStart ReadyMix	17 µL
	Illumina primer mix	2 µL
	Adaptor-ligated library	15 µL

57 Run the following cycles on a thermocycler:

	Cycles	Temperature	Time
	1	98°C	45 seconds
	12	98°C	15 seconds
		60°C	30 seconds
		72°C	30 seconds
	1	72°C	1 minute
		4°C	∞

58 Proceed immediately to cleanup or store at 4°C.

Post amplification cleanup

59 Allow Mag-Bind TotalPure NGS beads to equilibrate to room temperature, vortex until homogenous.

60 Add 27.2 µL of beads to 34 µL of amplified product, mix well, and incubate at RT for 10 minutes.

61 Place tubes on a magnetic stand and incubate until solution appears clear.

62 Discard supernatant without disturbing the beads.

63 While tubes are on the magnet, add 200 µL of 80% EtOH, incubate for 30 seconds, and discard the EtOH wash.

64 Repeat previous 80% EtOH wash and remove as much EtOH as possible.

65 Leave tubes on magnet and air dry for 5 minutes.

66 Remove tubes from magnet and add 25 µL of Tris-EDTA or elution buffer. Mix well by pipetting.

67 Place tubes on magnet stand. When solution appears clear, remove supernatant without disturbing the beads and place into new tubes.

68 *Note: If your lab has a KingFisher, you can download our automated protocols here:* https://github.com/grubaughlab/Kingfisher_protocols (use 'purification.bdz' for this step)

Library quantification and pooling

69 Quantify the DNA concentration of each sample (1 µL) using the Qubit High Sensitivity DNA kit.

70 Pool equal concentrations (e.g., 1-10 ng) of each library for sequencing.

71 Check DNA fragment distributions of the pooled sample using the BioAnalyzer DNA 1000 kit. Peak fragment size from 400 bp tiled amplicons with proper ligated adaptors should

be ~ 580 nt. If ~180 bp bands (adaptor dimers) still exist, perform post amplification cleanup again.

72 Quantify the DNA concentration of the pooled library (1 µL) using the Qubit High Sensitivity DNA kit.

73 Note: At least 0.76 ng/µL is required to achieve 2 nM for library pooling. Libraries will need to be concentrated or re-amplified if less than this amount.

74 **Convert DNA libraries from weight to moles:**
Molecular weight [nM] = Library concentration [ng/µL] / ((ave. library size x 650)/1,000,000)
Example: if ave. size of library is 580 bp and concentration is 2.5 ng/µL...

```
(580 x 650) / 1,000,000 = 0.377
5 / 0.377 = 6.6 nM
```

75 Dilute the pooled library to 2 nM in 10 mM TE.

76 (Optional) Ensure the library molar concentration using the Kapa Library Quantification kit.

77 **If sending your sample to a genomics core (i.e., not loading the MiSeq yourself), stop here.**

Diluting the pooled library for sequencing

78 Combine 10 µL of the 2nM pooled library to 10 µL of 0.1 N NaOH and mix. Incubate from 5 minutes at room temperature to denature the dsDNA.

79 Add 980 µL of HT1 (comes with the MiSeq kits). New concentration = 20 pM.

80 Dilute to the desired concentration using the following volumes.

	Concentration	10 pM	12 pM	14 pM	16 pM

	20 pM Library	295 µl	255 µl	415 µl	475 µl
	Prechilled HT1	300 µl	240 µl	180 µl	120 µl
	*PhiX control **	5 µl	5 µl	5 µl	5 µl

*PhiX control should also be denatured and diluted to 20 pM.

81 Note: loading too high of a sample on a MiSeq leads to over-clustering and decreased quality, which may make the data unusable. Adding too low leads to under-clustering and may not generate enough data for sufficient sequencing coverage. In our hands, optimal cluster densities were reached using 10-12 pM with the MiSeq v2 kits and 14-16 pM with the MiSeq v3 kits. Loading concentrations should be empirically determined with each lab.

82 Following loading instructions located in the MiSeq user guides.

Data processing and analysis

83 Use iVar, following instructions on: **www.github.com/andersen-lab/ivar**.

An example pipeline for generating consensus sequences and utilizing barcode transcripts to estimate contamination can be found on: **www.github.com/watronfire/PrimalSeq_Pipeline**