

The basics of DNA barcoding using the MinION

Stefan Prost^{1,2,3,4}

¹Natural History Museum Vienna, Central Research Laboratories, Vienna, Austria

²VetmedUni, Konrad Lorenz Institute of Ethology, Vienna, Austria

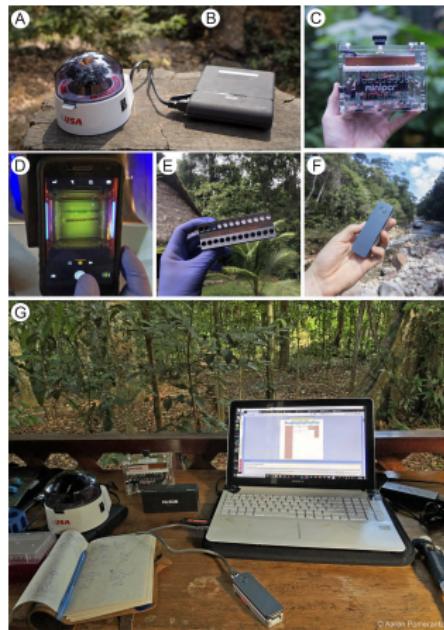
³University of Vienna, Department of Cognitive Biology, Vienna, Austria

⁴SANBI National Zoological Garden, Pretoria, South Africa

November 12th, 2021



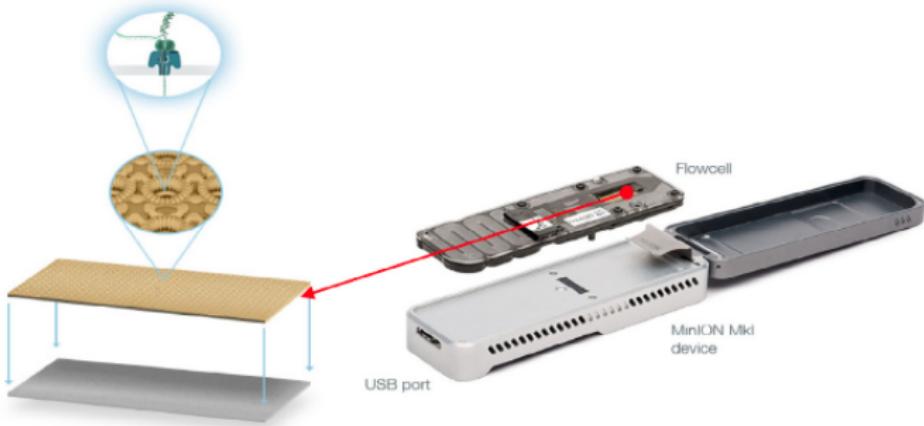
- ① MinION technology
- ② PCR setup
- ③ Multiplexing
- ④ Library Prep
- ⑤ Sequencing



Poermantz et al. (*in prep*)

- ① MinION technology
- ② PCR setup
- ③ Multiplexing
- ④ Library Prep
- ⑤ Sequencing





Lu, Giaordano and Ning, 2016 *Genomics Proteomics Bioinformatics*

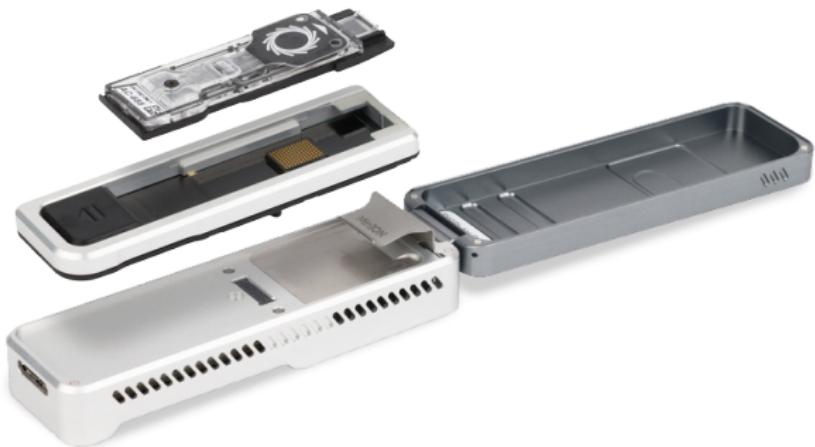


The Flongle Flow Cell can generate up to 2.8 Gb of data enabling direct, real-time DNA & cDNA sequencing on smaller, single-use flow cells.

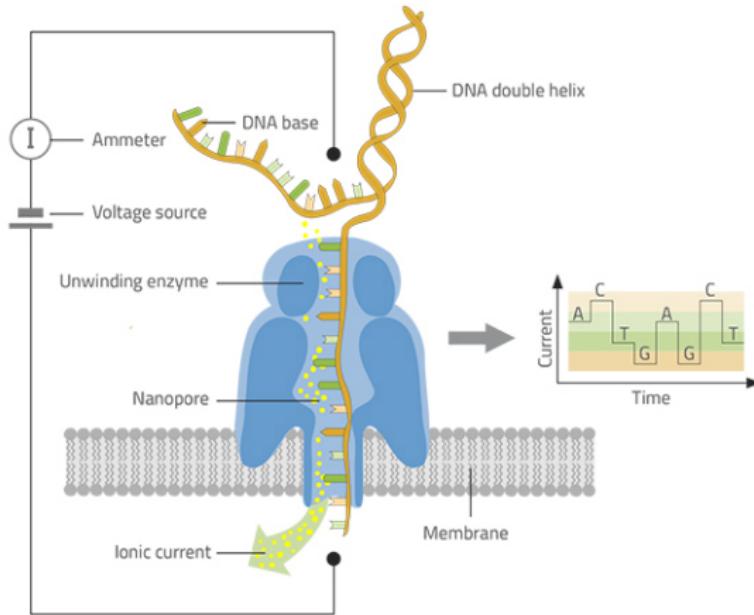


The MinION Flow Cell can generate up to 50 Gb of data for sequencing DNA, cDNA or native RNA in real-time.

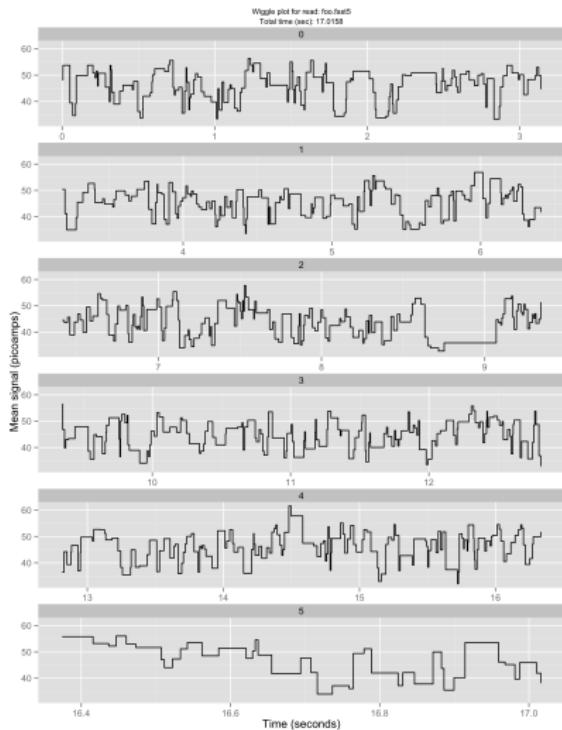
Oxford Nanopore Technologies

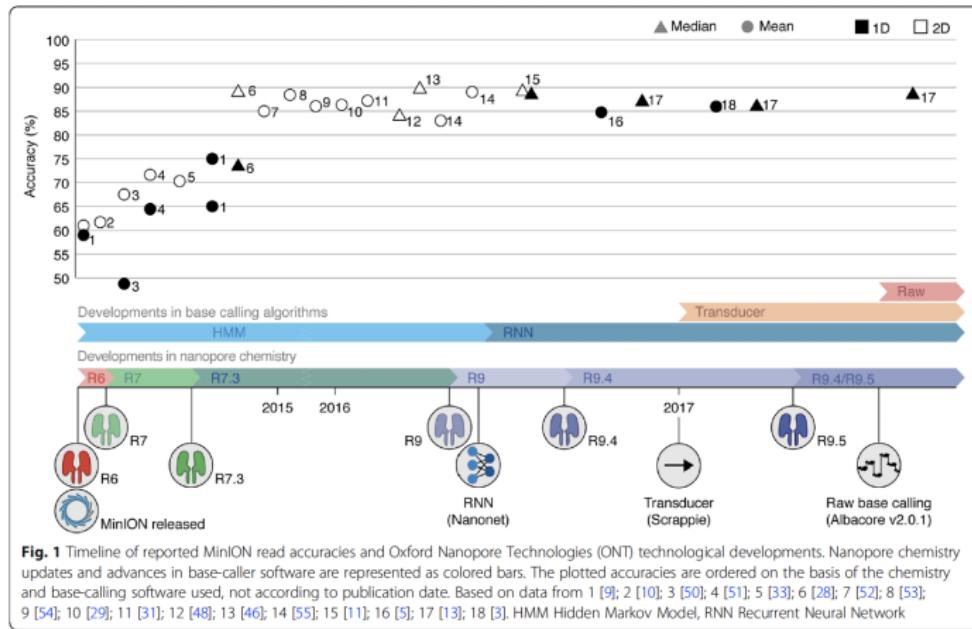


Oxford Nanopore Technologies



Kerstin Goepfrich (www.scienceinschool.org)





Rang et al. 2014 *Genome Biology*

R10: a new nanopore



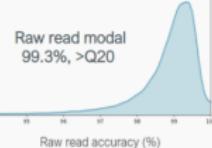
Oxford Nanopore Technologies

Q20+ Sequencing kit

Run today on MiniON / GridION



- With fuel fix, *high* capture adapter
- Q20+ Motor
- Includes all other ligation sequencing kit components



Duplex enabled



Duplex pairs template and complement strands
for highest accuracy data



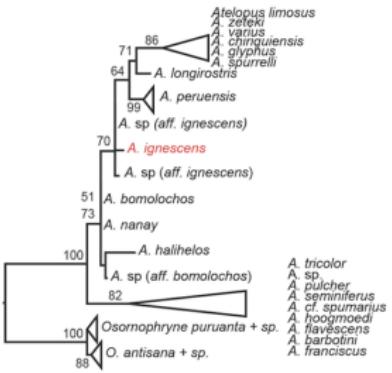
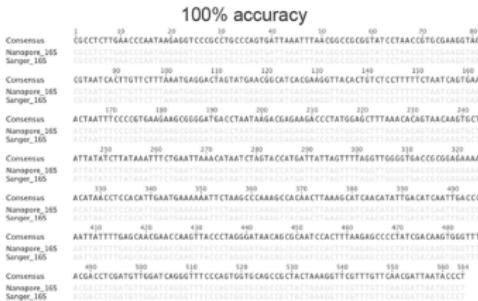
Early Access underway

Oxford Nanopore Technologies

C



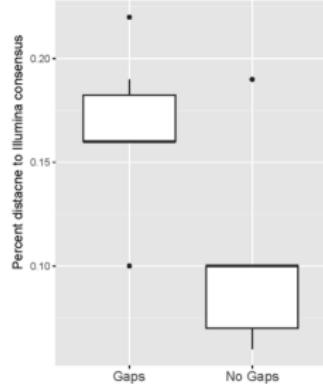
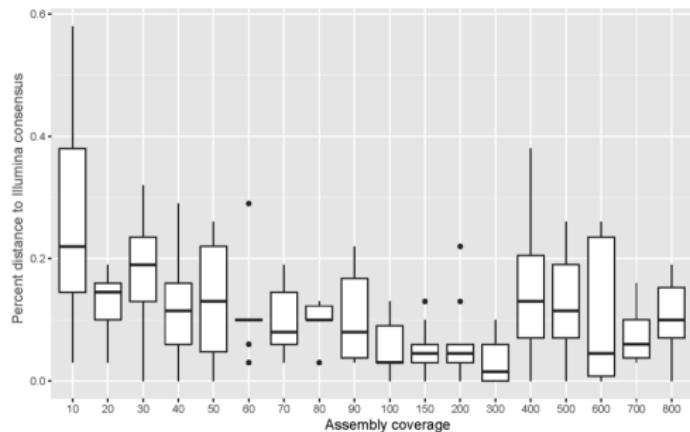
Atelopus ignescens



0.02

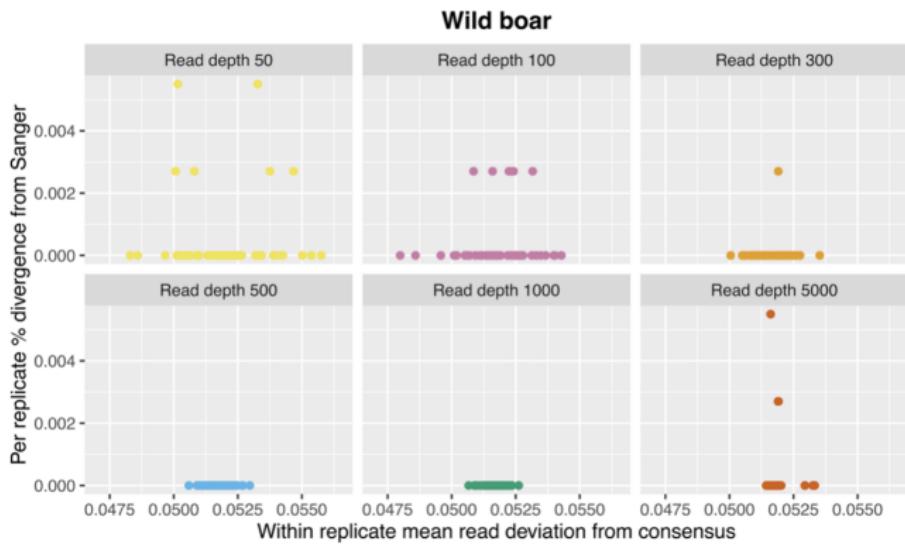
Pomerantz et al. 2018 *GigaScience*

Consensus Accuracy



Krehenwinkel et al. 2019 *GigaScience*

Consensus Accuracy



Vasiljevic et al. 2021 *Forensic Science International Genetics*



July 2017 Ecuador, Pomerantz et al. 2018 *GigaScience*
Reviewed in Krehenwinkel, Pomerantz and Prost 2020 *Genes*

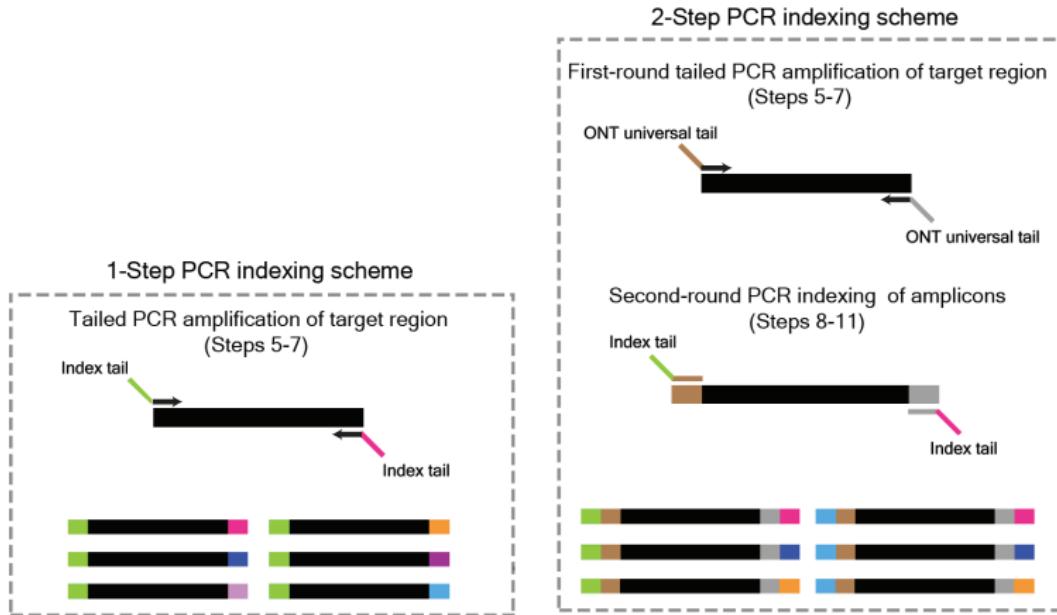


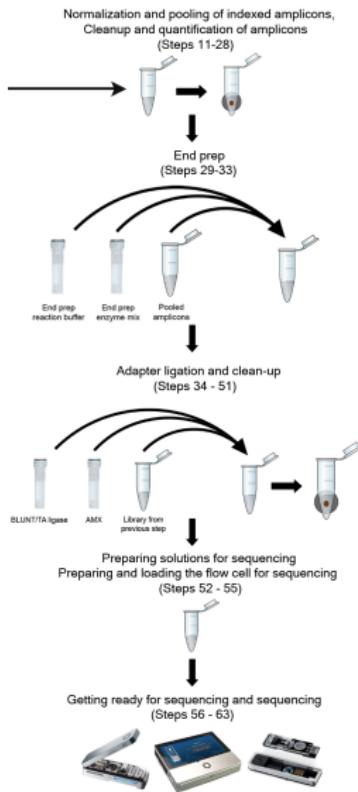
July 2017 Ecuador, Pomerantz et al. 2018 *GigaScience*

- ① MinION technology
- ② PCR setup
- ③ Multiplexing
- ④ Library Prep
- ⑤ Sequencing



Poermantz et al. (*in prep*)

Pomerantz et al. *in prep*

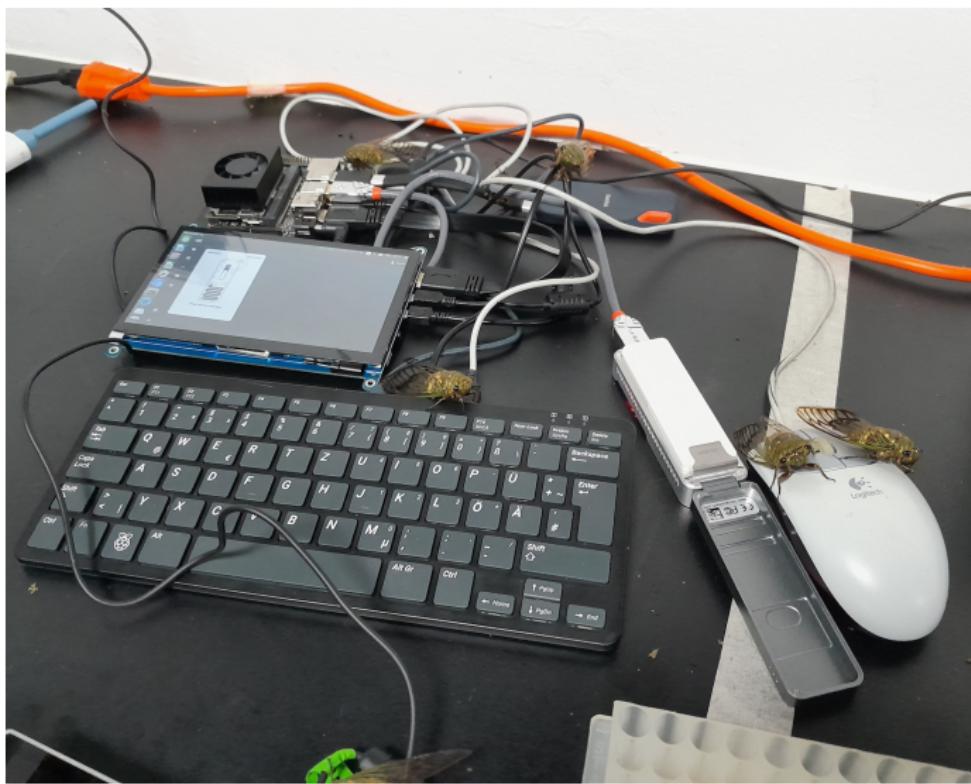


July 2017 Ecuador, Pomerantz et al. 2018
GigaScience

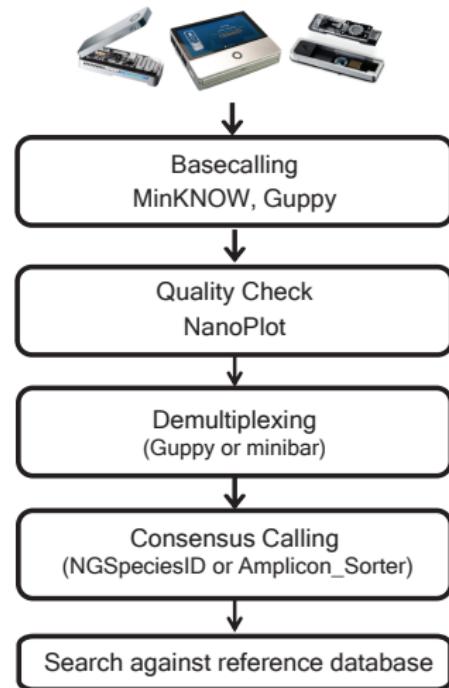
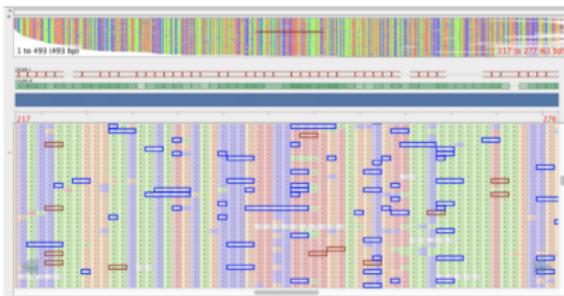
Pomerantz et al. *in prep*



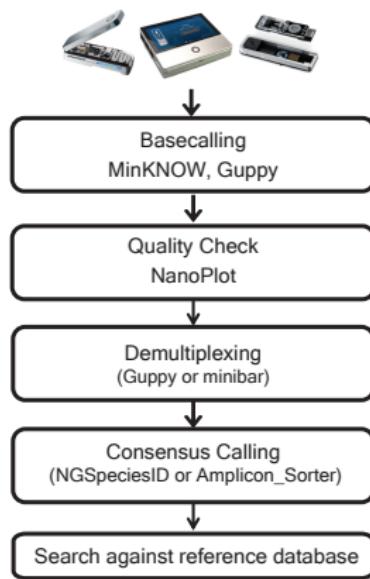
July 2017 Ecuador, Pomerantz et al. 2018 *GigaScience*



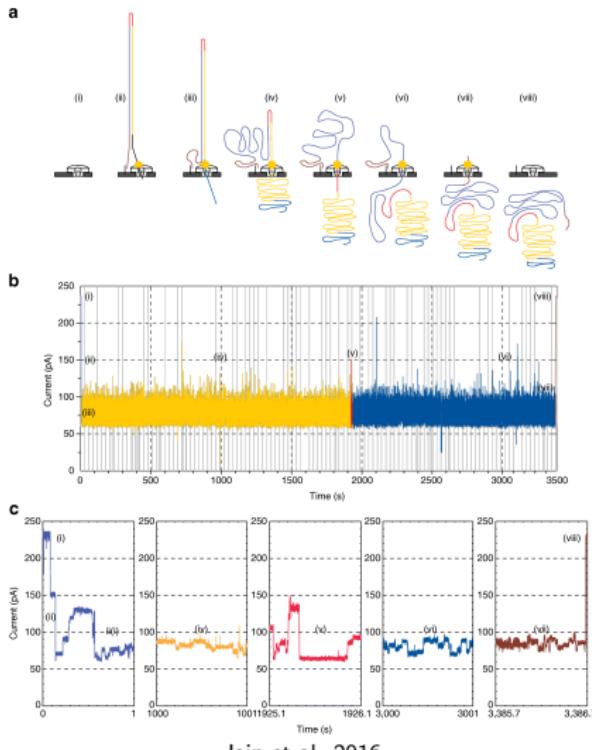
Peru 2021



- ① Base-calling
- ② De-multiplexing
- ③ Raw read quality assessment
- ④ Consensus generation
- ⑤ Error-correcting
- ⑥ Species ID

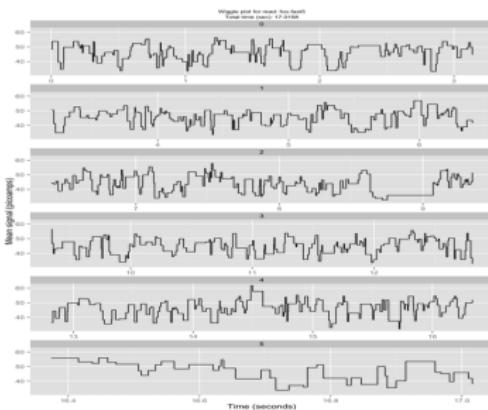


- ① Base-calling (Guppy)
- ② De-multiplexing (Guppy, qcat or Minibar)
- ③ Raw read quality assessment (Nanoplot)
- ④ Consensus generation (NGSpeciesID)
- ⑤ Error-correcting (Racon)
- ⑥ Species ID (Blast search)



Jain et al. 2016

- ① Base-calling (Guppy)
- ② De-multiplexing (Guppy, qcat or Minibar)
- ③ Raw read quality assessment (Nanoplot)
- ④ Consensus generation (NGSpeciesID)
- ⑤ Error-correcting (Racon)
- ⑥ Species ID (Blast search)

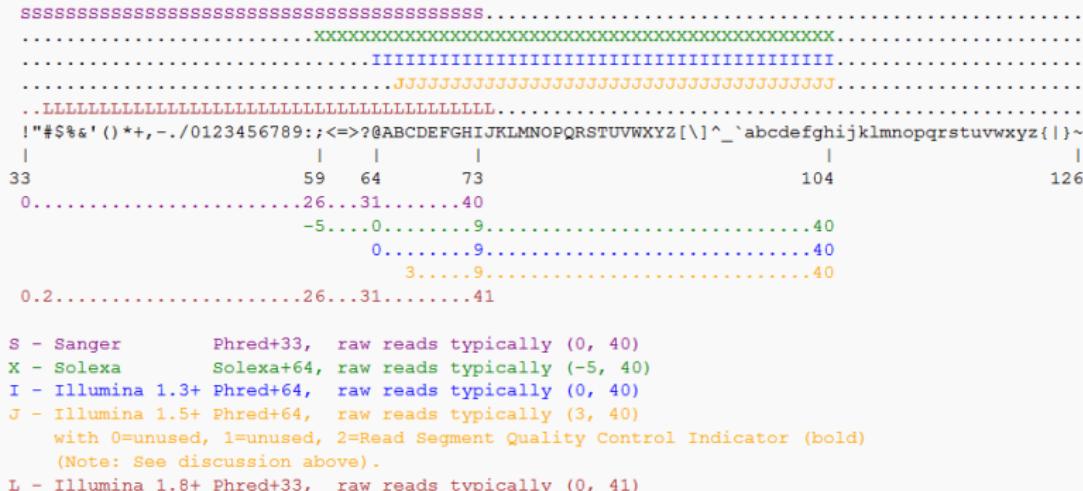


```
>read_1
TTTATAACGCTCACCCCTCTGCATCTC
```

Fastq Format

```
@HISEQ:119:C42B3ACXX:7:1101:2009:2249 2:N:0:CGACCTG
TCTTGGGGACAGGGAAATTCAATTCAAATGAAAATCCTCAAAGAACGCC
TTTATTTACAGGAGGCTGTATATCTTAGCCAAAGTGGTAGATCGGAAGA
+
BB<BFFFFBFB<BFF7BBF<BF<FBB<FFF<FFBF<BFFFFFBFBF<7BB<BFBFB<BFFFFFF<FFFFFFF<BBFB<BB<BBBBBBBB7<B<BB<77<BBB7
```

Quality Score Encoding



- ① Base-calling (Guppy)
- ② De-multiplexing (Guppy, qcat or Minibar)
- ③ Raw read quality assessment (Nanoplot)
- ④ Consensus generation (NGSpeciesID)
- ⑤ Error-correcting (Racon)
- ⑥ Species ID (Blast search)

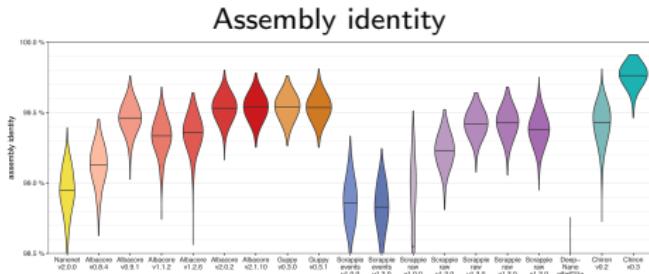
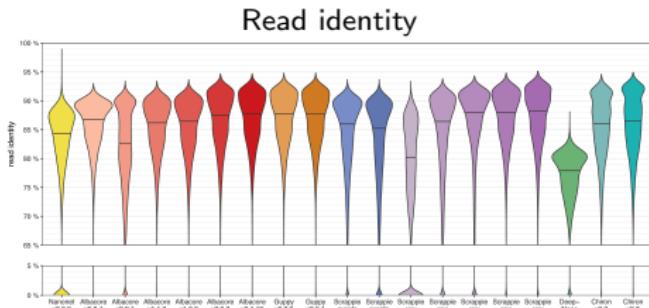


Guppy Basecaller



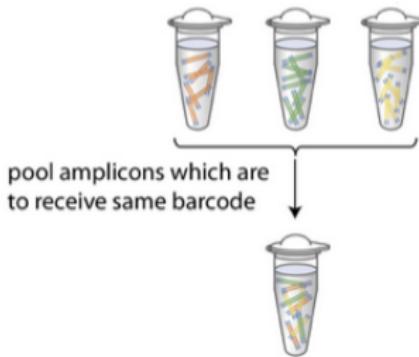
Bonito Basecaller

- ① Base-calling (Guppy)
 - ② De-multiplexing (Guppy, qcat or Minibar)
 - ③ Raw read quality assessment (Nanoplot)
 - ④ Consensus generation (NGSpeciesID)
 - ⑤ Error-correcting (Racon)
 - ⑥ Species ID (Blast search)

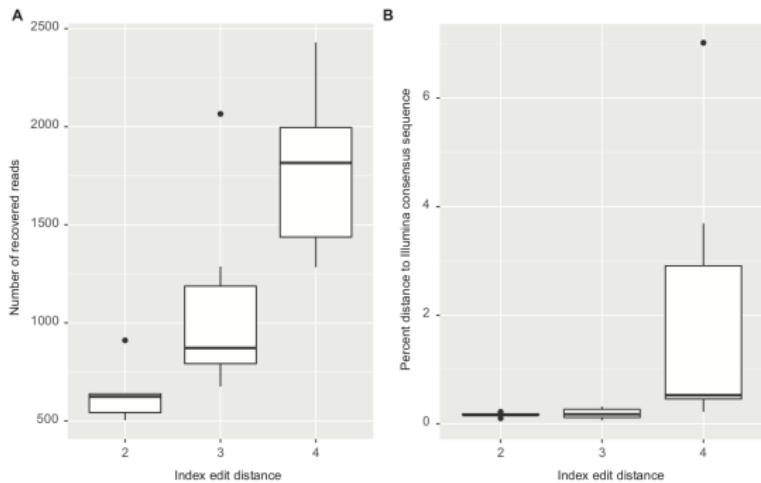


<https://github.com/rrwick/Basecalling-comparison>

- ① Base-calling (Guppy)
- ② De-multiplexing (Guppy, qcat or Minibar)
- ③ Raw read quality assessment (Nanoplot)
- ④ Consensus generation (NGSpeciesID)
- ⑤ Error-correcting (Racon)
- ⑥ Species ID (Blast search)

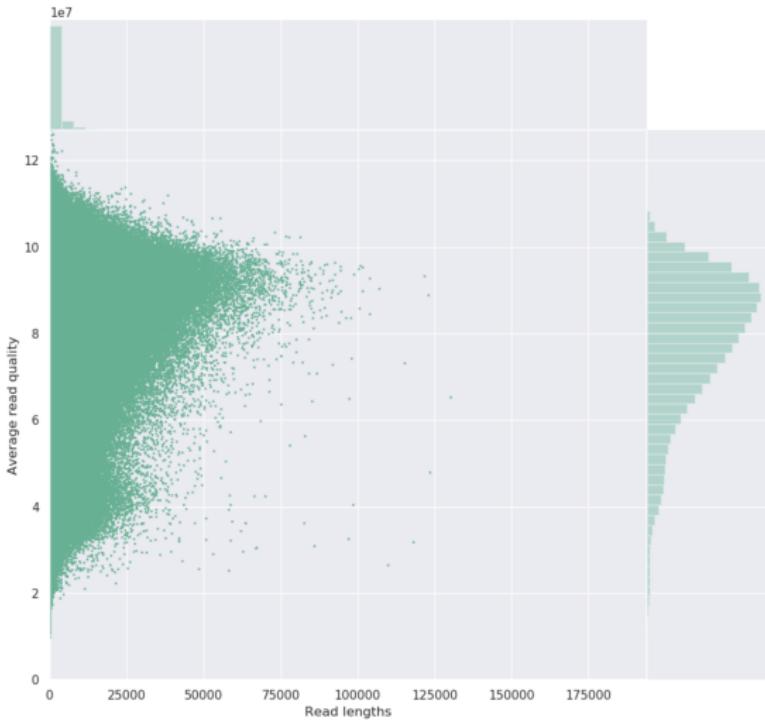


- ① Base-calling (Guppy)
- ② De-multiplexing (Guppy, qcat or Minibar)
- ③ Raw read quality assessment (Nanoplot)
- ④ Consensus generation (NGSpeciesID)
- ⑤ Error-correcting (Racon)
- ⑥ Species ID (Blast search)

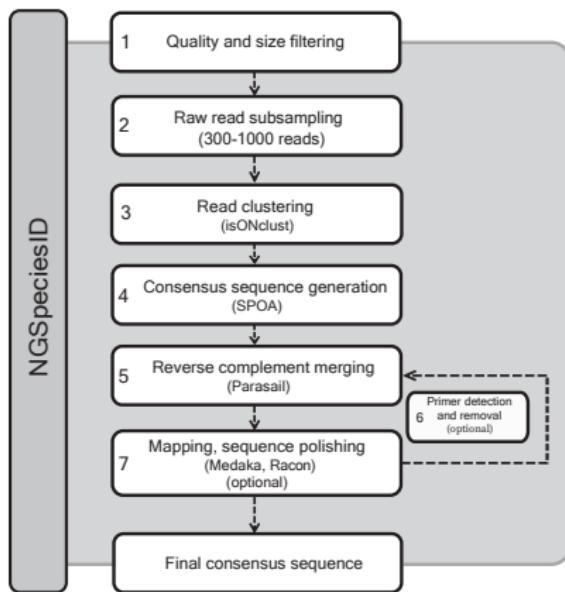


Krehenwinkel et al. 2019

- ① Base-calling (Guppy)
- ② De-multiplexing (Guppy, qcat or Minibar)
- ③ Raw read quality assessment (Nanoplot)
- ④ Consensus generation (NGSpeciesID)
- ⑤ Error-correcting (Racon)
- ⑥ Species ID (Blast search)

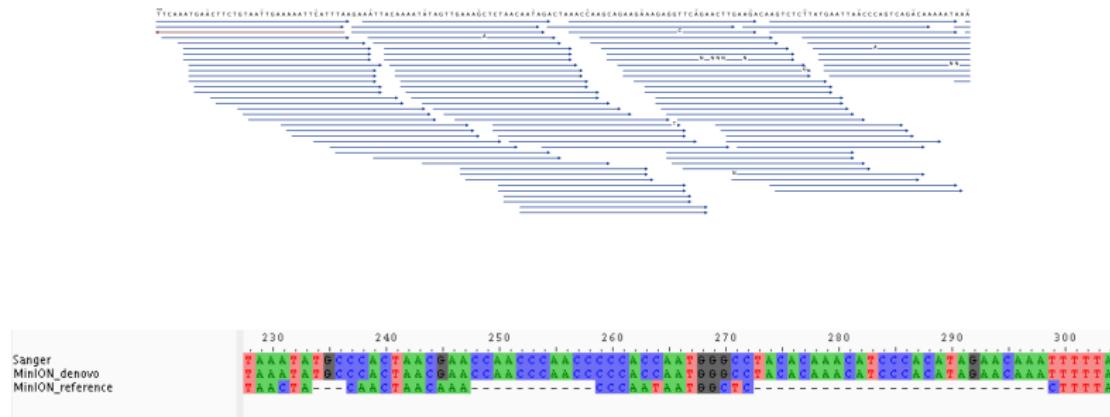


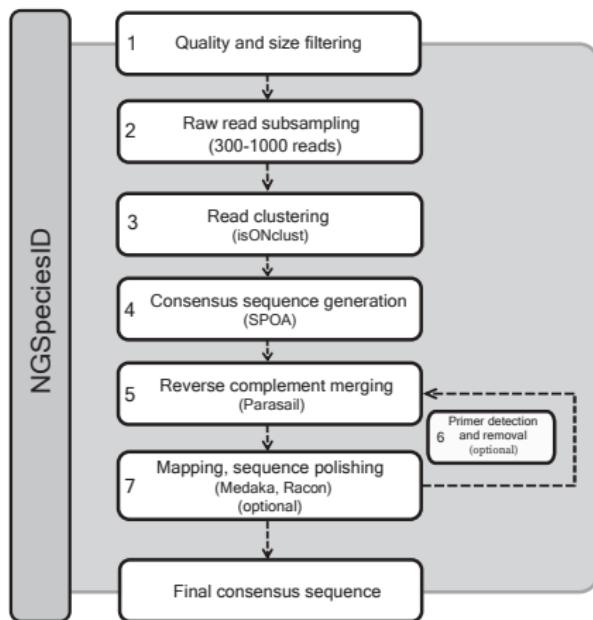
- ① Base-calling (Guppy)
- ② De-multiplexing (Guppy, qcat or Minibar)
- ③ Raw read quality assessment (Nanoplot)
- ④ Consensus generation (NGSpeciesID)
- ⑤ Error-correcting (Racon)
- ⑥ Species ID (Blast search)



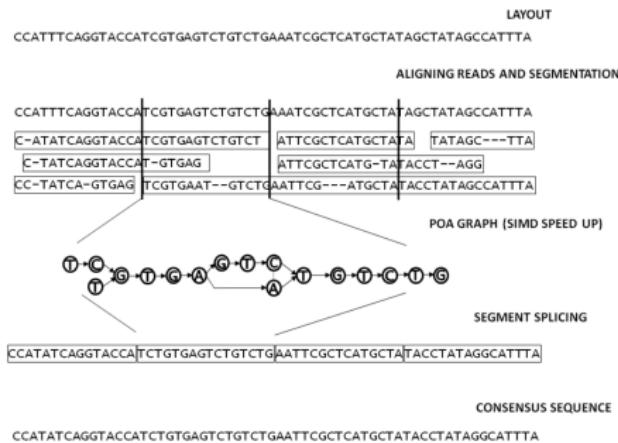
Sahlin, Lim and Prost 2020

Reference-based mapping





- ① Base-calling (Guppy)
- ② De-multiplexing (Guppy, qcat or Minibar)
- ③ Raw read quality assessment (Nanoplot)
- ④ Consensus generation (NGSpeciesID)
- ⑤ Error-correcting (Racon)
- ⑥ Species ID (Blast search)



Vaser et al. 2016

⑥ Species ID (Blast search)

RID: KBK75XD6014 (Expires on 07-23 22:49 pm)

Query ID: kcl|Query_33635
Description: filtered_reads_rsc2
Molecule type: dna
Query Length: 958

Database Name: nr
Description: Nucleotide collection (nt)
Program: BLASTN 2.9.0+ >[Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [MSA viewer](#)

[Graphic Summary](#)

Distribution of the top 100 Blast Hits on 100 subject sequences. [View](#)
 Mouse over to see the file, click to show alignment.

Color key for alignment scores:
 ■ <40 ■ 40-50 ■ 50-80 ■ 80-200 ■ >=200

Query

1 150 300 450 600 750 900

[Descriptions](#)

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

[Alignments](#) [Distance matrix](#) [Taxonomy](#) [Distance tree of results](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
1	Alouatta seniculus mitochondrion, complete genome	763	763	72%	0.0	86.60%	HQ644333.1
2	Alouatta seniculus mitochondrion, complete genome	747	747	72%	0.0	86.20%	KY202428.1
3	Atelopus bezebuth mitochondrion, complete genome	730	730	72%	0.0	85.75%	KC757396.1
4	Alouatta caraya mitochondrion, complete genome	730	730	72%	0.0	85.77%	KC757384.1
5	Atelopus bezebuth mitochondrion, complete genome	730	730	72%	0.0	85.75%	FJ75422.1
6	Atelopus geoffroyi mitochondrial gene for cytochrome C oxidase subunit I, complete cds	725	725	72%	0.0	85.61%	AB016730.1

- ① Base-calling (Guppy)
- ② De-multiplexing (Guppy, qcat or Minibar)
- ③ Raw read quality assessment (Nanoplot)
- ④ Consensus generation (NGSpeciesID)
- ⑤ Error-correcting (Racon)
- ⑥ Species ID (Blast search)

- **Bit Score**

- Measures the quality of the alignment.

- **Query Coverage**

- How much of the query matches.

- **E-value**

- Likelihood that a given sequence match is purely by chance (depends on the database size)

- **Percent Identity**

- How similar are the query and the database sequence.

Stefan Prost

stefanprost.research@protonmail.com

Photo by: Tropical Herping (Ecuador)



stefan.prost@nhm-wien.ac.at

