Increasing the accuracy of nanopore DNA sequencing using a time-varying cross membrane voltage

Matthew T. Noakes, Henry Brinkerhoff, Andrew H. Laszlo, Ian M. Derrington, Kyle W. Langford, Jonathan W. Mount, Jasmine L. Bowman, Katherine S. Baker, Kenji M. Doering, Benjamin I. Tickman and Jens H. Gundlach

Nanopore DNA sequencing is limited by low base-calling accuracy. Improved base-calling accuracy has so far relied on specialized base-calling algorithms, different nanopores and motor enzymes, or biochemical methods to re-read DNA molecules. Two primary error modes hamper sequencing accuracy: enzyme mis-steps and sequences with indistinguishable signals. We vary the driving voltage from 100 to 200 mV, with a frequency of 200 Hz, across a Mycobacterium smegmatis porin A (MspA) nanopore, thus changing how the DNA strand moves through the nanopore. A DNA helicase moves the DNA through the nanopore in discrete steps, and the variable voltage moves the DNA continuously between these steps. The electronic signal produced with variable voltage is used to overcome the primary error modes in base calling. We found that single-passage de novo base-calling accuracy of 62.7 ± 0.5% with a constant driving voltage improves to 79.3 ± 0.3% with a variable driving voltage. The variable-voltage sequencing mode is complementary to other methods to boost the accuracy of nanopore sequencing and could be incorporated into any enzyme-actuated nanopore sequencing device.
when the enzyme backtracks to a position along the DNA observed previously (Fig. 1b, blue stars). The existence of irregular enzyme steps means that the observed time order of conductance states does not necessarily match the base order in the DNA, and causes sequencing errors.

We reasoned that replacing the constant bias voltage with a time-varying voltage would reduce the impact of both of these error modes substantially. In our sequencing experiments, we used Mycobacterium smegmatis porin A (MspA) as our nanopore. MspA has a single narrow constriction region that is ideally suited to resolve nucleotide-long enzyme steps along single-stranded DNA (ssDNA)\(^{24,25}\). We used the Hel308 DNA helicase enzyme from Thermococcus gammatolerans EJ3 (hereafter referred to as Hel308) as the motor enzyme to control DNA translocation through the pore. Hel308 has been observed to take two steps per nucleotide as it translocates along ssDNA, with each step approximately half a nucleotide in length\(^{26}\). These half-nucleotide steps provide two conductance measurements per nucleotide (Fig. 1c).

Positive voltage applied across the nanopore generates a force on the DNA threaded through the pore. Variation of the magnitude of this voltage changes the force pulling on the DNA. The force stretches the section of DNA between the DNA-binding sites within Hel308 and the high field region at the nanopore’s constriction\(^{27}\). An increase in the applied voltage elongates the DNA and shifts the relative position of the DNA in the constriction (Fig. 1d). A voltage change from 100 mV to 200 mV repositions the DNA in the pore by slightly more than a full nucleotide (Fig. 1e). The applied voltage serves as a fine control over the DNA position in the pore.

The fine DNA position control that is achieved using the variable voltage complements the discrete stepping of the motor enzyme. We combine the enzyme and voltage control methods by replacing the constant applied voltage with a 200 Hz, symmetrical triangle waveform from 100 to 200 mV. The positive overall bias is necessary to hold the DNA–enzyme complex on top of the pore. The 200 Hz triangle wave frequency goes through several cycles for each Hel308 step (average rate \(\sim\) 20 steps s\(^{-1}\) in our sequencing conditions, Supplementary Note 2). While the motor enzyme steps along the entire length of the DNA, the changing voltage repositions the...
DNA incrementally within each enzyme step. Together, the enzyme steps and the variable voltage sample the effect of the DNA on the pore’s conductance nearly continuously along the DNA (Fig. 1f). In the constant-voltage signal, the pore conductance is probed only at a single DNA position at each enzyme step. Each step is thus characterized by only a single value: the conductance at that DNA position (Fig. 2a,b, top). Variable-voltage sequencing instead probes the conductance continuously over an approximately one-nucleotide-long range at each enzyme step, characterizing each step by a conductance-versus-position curve (Fig. 2a,b, bottom). These curve segments provide additional identifying information as to the generating DNA sequence compared to the mean conductance alone (Supplementary Note 3). Two sequences with nearly identical conductance values in the constant-voltage mode can be distinguished based on the shape of the curves generated by the variable voltage (Fig. 2a, orange highlights).

The variable-voltage signal also provides information about the correct ordering of the measured enzyme steps and can be used to infer the existence of steps that are too fast to observe. The ability of the variable-voltage technique to identify enzyme mis-steps is enabled by the continuous sampling of the conductance through the pore as a function of DNA position. In variable-voltage sequencing, at each consecutive Hel308 half-nucleotide step, the full-nucleotide stretch caused by the voltage sweep samples the conductance at many of the same DNA positions as the previous and next Hel308 steps. Therefore, each measured segment of the conductance-versus-position curve will be overlapping and continuous with the segments measured at adjacent Hel308 steps (Supplementary Note 4). If two consecutively measured segments are not overlapping and continuous, a non-uniform step such as a backstep or skip must have occurred. The degree of overlap between consecutive measurements can therefore be used to identify and correct enzyme mis-steps.
mis-steps (Fig. 2a,b, colored arrows). A probabilistic support vector machine informed by the shapes of the curves immediately preceding and following each enzyme step (Supplementary Note 5) is used on the variable-voltage signal to identify and eliminate misorderings caused by enzyme mis-steps and re-establish the order that is most representative of the generating DNA sequence. The resulting corrected signal (Fig. 2c,d) is free of enzyme mis-steps and is more easily decoded into the correct DNA sequence. Measurements of DNA positions that go completely unobserved owing to enzyme skips cannot be filled in at this stage. However, the overlap information can be used to label the probable locations of enzyme skips in the final signal to be sequenced. This information tells the sequencer that one or more bases must be inserted at this location, reducing the detrimental impact of enzyme skips relative to constant-voltage nanopore sequencing.

To evaluate objectively the extent to which the variable-voltage method improves single-passage sequencing accuracy over the constant-voltage method, we tested both sequencing methods on the same target DNA sequence, using the same enzyme (Hel308) and nanopore (MspA). In both cases, we used a hidden Markov model (HMM) (Supplementary Note 6) to decode the generating DNA sequence for the observed signal. For both constant- and variable-voltage sequencing, we used a model mapping each unique six-base sequence segment (6-mer) to an associated conductance signal. We generated this model empirically by measuring the signal of known DNA sequences (ΦX-174 and lambda phage DNA, as well as synthetic oligonucleotides), using our variable-voltage sequencing conditions (Supplementary Note 7). For constant-voltage sequencing, we extracted a constant-voltage 6-mer model from the variable-voltage model (Supplementary Note 8) to ensure that any systematic

---

**Fig. 3 | Performance using constant- and variable-voltage sequencing.** a–d, Confusion matrices for sequencing using constant-voltage (a) and variable-voltage (b) show a reduction in mismatch, insertion (Ins), and deletion (Del) errors across all bases with the use of variable voltage. Histograms of single-passage accuracies for 31 constant-voltage reads (9,368 total bases; 2,203 called As, 2,188 called Cs, 2,166 called Gs, 2,144 called Ts, 667 called gaps) (c) and 97 variable-voltage reads (17,309 total bases; 4,021 called As, 4,073 called Cs, 3,813 called Gs, 4,081 called Ts, 1,321 called gaps) (d) show a significant improvement in the distribution of read accuracies using variable voltage. The distribution of of single-read accuracies for the 31 constant-voltage reads (c) and 97 variable-voltage reads (d) are plotted in blue, with the distribution of accuracies for randomly generated sequences of the same lengths plotted in red. Although constant-voltage nanopore sequencing is only a few percent above random base calling, the variable-voltage method yields a substantial improvement.
model errors affecting the sequencing accuracy of one method affected both methods equally.

We used the pET28a vector as the target DNA sequence because it provided a non-synthetic DNA testing ground for the two methods separate from the sequences that were used in constructing the 6-mer model. We fragmented the pET28a vector using a selection of restriction enzymes (Supplementary Note 9), enabling the necessary sequencing adapters to be attached and increasing the likelihood of reading sequences at all locations on the pET28a vector given the limited processivity of Hel308 (∼1,000 nucleotides). The variable-voltage method does not reduce the ability of the nanopore sequencer to sequence long (multiple kilobase) reads (Supplementary Note 10).

We obtained reads of plasmid pET28a fragments, using both constant-voltage (31 reads, \( n = 9,368 \) bases across all reads) and variable-voltage (97 reads, \( n = 17,309 \) bases across all reads) methods. Enzyme steps in the constant- and variable-voltage conductance signals were detected using a custom change-point detection algorithm (Supplementary Note 11), segmenting the data into distinct conductance states. In the variable-voltage experiments, the capacitive charging currents from the bilayer were removed from each state by using custom software (Supplementary Note 12). We used the overlap information between successive conductance states to identify and correct enzyme mis-steps in the variable-voltage reads (Supplementary Note 5), and then both sets of reads were calibrated and sequenced. For both the constant- and variable-voltage sequencing results, we determined the ground truth sequence for each read by aligning the called sequence to the pET28a reference sequence. According to the alignment, we calculated the per-base sequencing accuracy as \( (\text{number of matches} + \text{number of deletions}) + \text{number of mismatches} + \text{number of insertions} + \text{number of deletions}) \) divided by \( N \), where \( N \) is the number of bases sequenced. Relative to the constant-voltage reads, the variable-voltage reads have fewer base-calling errors (mismatches, deletions, and insertions; Fig. 3a,b). The average per-base accuracy of the variable-voltage reads is \( 79.3\pm0.3\% \) (s.e.m.) for single passages of a single-stranded DNA molecule. This represents a substantial improvement compared with nanopore sequencing using constant voltage, which had an average accuracy of \( 62.7\pm0.5\% \) (s.e.m.) for the same DNA sample. Our constant-voltage sequencing accuracies are similar to single-passage, unpolished 1D reads reported for the Oxford Nanopore Technologies MinION device\(^{26}\). To contextualize the relative accuracy of the two methods, we compared the distribution of observed per-read accuracies with the accuracy distribution for random sequences of the same lengths aligned against the pET28a reference sequence. The sequencing accuracies of these random sequences is approximately 58\% (Fig. 3c,d); this random base-call accuracy is so high (that is, much higher than 25\%) because of the freedom provided to the alignment algorithm to call insertions, deletions, or mismatches. The constant-voltage read accuracies barely outperform the accuracies of randomly generated sequences (Fig. 3c), whereas the variable-voltage read accuracies are substantially higher than the distribution of random accuracies (Fig. 3d).

We conclude that the variable-voltage method recovers significantly more information from the target DNA and thereby substantially increases the base-calling accuracy. Improved single-read accuracy should enable fewer reads to be assembled into a high-accuracy consensus sequence, thereby reducing sequencing time and cost. In addition, variable-voltage sequencing overcomes systematic errors, such as sequence-dependent enzyme mis-steps\(^{26}\) and indistinguishable signals, which persist even when the information from many reads is combined. Variable-voltage reads can be more confidently identified with only single-read coverage. This capability is necessary for nanopore sequencing applications in which high coverage is not an option, such as metagenomics studies or pathogen detection at low concentrations.

The additional information provided by the variable-voltage signal is complementary to other nanopore sequencing improvements, including more processive and predictable motor enzymes, more sophisticated base-calling algorithms\(^{11–13}\), reading both sense and antisense of the target DNA strand in ‘2D’ techniques\(^{14}\) (or the similar ‘1D squared’ technique), or polishing reads with a consensus of passages of different DNA molecules\(^{15,35}\). Existing nanopore sequencers already consist of hundreds or thousands of parallel nanopores that can be addressed separately with distinct driving voltages, so the variable-voltage method requires little re-engineering of the nanopore sequencing device other than the application of a waveform in place of a constant voltage. Consequently, our method could be used to improve sequencing accuracy of most existing platforms. The performance of variable-voltage nanopore sequencing will improve further as larger data sets are used to train both the model that maps conductance curves to DNA bases and the enzyme mis-step classifier.

We anticipate that incorporating our variable-voltage method into nanopore sequencing platforms will enable wide-scale improvement of all nanopore-based DNA sequencing applications, including species identification, epigenetic mapping, and higher accuracy de novo genome sequencing at lower coverage.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information, details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-019-0096-0.

Received: 28 September 2018; Accepted: 11 March 2019; Published online: 22 April 2019

References

1. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. Proc. Natl Acad. Sci. 93, 13770–13773 (1996).

2. Mannaro, E. A. et al. Reading DNA at single-nucleotide resolution with a mutantMspA nanopore and ph29 DNA polymerase. Nat. Biotechnol. 30, 349–353 (2012).

3. Laszlo, A. H. et al. Decoding long nanopore sequencing reads of natural DNA. Nat. Biotechnol. 32, 829–833 (2014).

4. Jain, M. et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat. Biotechnol. 36, 338–345 (2018).

5. Craig, J. M. et al. Direct detection of unnatural DNA nucleotides dNaM and d5SICS using the MspA nanopore. PLoS ONE 10, e0143253 (2015).

6. Laszlo, A. H. et al. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. Proc. Natl Acad. Sci. 110, 18904–18909 (2013).

7. Schreiber, J. et al. Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. Proc. Natl Acad. Sci. USA 110, 18910–18915 (2013).

8. Lindblad, A. C. et al. Mapping DNA methylation with high-throughput nanopore sequencing. Nat. Methods 14, 411–413 (2017).

9. Branton, D. et al. The potential and challenges of nanopore sequencing. Nat. Biotechnol. 26, 1146–1153 (2008).

10. Castro-Wallace, S. L. et al. Nanopore DNA sequencing and genome assembly on the International Space Station. Sci. Rep. 7, 18022 (2017).

11. Loman, N. J., Quick, J. & Simpson, J. T.: A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat. Methods 12, 733–735 (2015).
12. Theuns, S. et al. Nanopore sequencing as a revolutionary diagnostic tool for porcine viral enteric disease complexes identifies porcine kobivirus as an important enteric virus. Sci. Rep. 8, 9830 (2018).

13. Bronzato Badial, A. et al. Nanopore sequencing as a surveillance tool for plant pathogens in plant and insect tissues. Plant Dis. 1648–1652 (2018).

14. Cao, M. D. et al. Streaming algorithms for identification pathogens and antibiotic resistance potential from real-time MinION™ sequencing. GigaScience 5, 1–16 (2016).

15. Simpson, J. T. et al. Detecting DNA cytosine methylation using nanopore sequencing. Nat. Methods 14, 407–410 (2017).

16. Quick, J. et al. Real-time, portable genome sequencing for Ebola surveillance. Nature 530, 228–232 (2016).

17. Faria, N. R. et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature 546, 406–410 (2017).

18. Karatzopoulos, L. E. et al. Assessment of Metagenomic MinION and Illumina sequencing as an approach for the recovery of whole genome sequences of chikungunya and dengue viruses directly from clinical samples. Preprint at https://doi.org/10.1101/355560 (2018).

19. Pomerantz, A. et al. Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. GigaScience 7, gky033 (2018).

20. Cherf, G. M. et al. Automated forward and reverse ratcheting of DNA in a nanopore at 5-A precision. Nat. Biotechnol. 30, 344–348 (2012).

21. Derrington, I. M. et al. Subangstrom single-molecule measurements of motor proteins using a nanopore. Nat. Biotechnol. 33, 1073–1075 (2015).

22. Jain, M. et al. MinION analysis and reference consortium: phase 2 data release and analysis of R9.0 chemistry. F1000Res. 6, 760 (2017).

23. Boža, V., Brejová, B. & Vinař, T. DeepNano: deep recurrent neural networks for base calling in MinION nanopore reads. PLoS ONE 12, e0178751 (2017).

24. Simpson, J. Signal-level algorithms for MinION data. Github https://github.com/js/nanopolish (2018).

Acknowledgements
This work was supported by the National Human Genome Research Institute grant R01HG005115 (to J.H.G.).

Competing interests
The authors J.H.G., M.T.N., and H.B., along with the University of Washington, have filed provisional patent applications covering the methods presented in this work. The patent has been filed under application number 62/805,870 by the University of Washington CoMotion.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41587-019-0096-0.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to J.H.G.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Methods

Proteins. The same mutant MspA protein was used in all sequencing experiments. This mutant, M2-NNN-MspA, was custom ordered from GenScript. M2-NNN-MspA is engineered on the wild-type MspA (accession number CAB56952.1) with the following mutations: D90N/D91N/D93N/D118R/E119K/D134R (ref. 36). All sequencing experiments used the Hel308 helicase enzyme from *T. gammatolerans* EJ3 (accession number WP_015858487.1). Hel308 was expressed in *Escherichia coli* by using standard techniques. All proteins were stored at −20 °C until immediately before use.

DNA sequences and constructs. Short DNA oligonucleotides were synthesized and purified using column purification methods at the Stanford University Protein and Nucleic Acid Facility. The *Φ*X−174 DNA (NCBI reference sequence NC_001422.1) was obtained from New England Biolabs. The lambda phage DNA (GenBank J02459.1) was obtained from Promega. The pET-28a DNA was obtained from collaborators who used it as an expression vector for another DNA sequence not used in this work. The complete DNA sequences for *Φ*X-174, lambda, and pET28a can be found at https://doi.org/10.6084/m9.figshare.7140896.v1.

All experiments were conducted using the DNA threaded through the pore 5′ end first. DNA constructs for Hel308 experiments consisted of a template read strand and a cholesterol-tagged blocking strand. A negatively charged terminal phosphate was attached to the 5′ end of the template strand, increasing the capture rate of that end by MspA. The cholesterol tag at the 5′ end of the blocking strand anchors the DNA constructs into the bilayer, increasing the local concentration near the pore and increasing the capture rate. A detailed description of how the various DNA sequences were prepared can be found in Supplementary Note 7. The set of custom DNA sequences used in our experiments is provided in Supplementary Note 13.

Nanopore experiments. All experiments were conducted as described previously36. In brief, experiments were established with a device made from Teflon that contains two ~50 μl chambers (cis and trans). The two chambers are connected by a Teflon heat-shrink ‘u-tube’, ~30 μl in volume. The cis side of the u-tube narrows into a horizontal ~20 μm aperture. Both chambers and the u-tube were filled with the operating buffers. The cis chamber was connected to ground via an Ag/AgCl electrode, while the trans-side Ag/AgCl electrode was connected to an Axopatch 200B integrating patch clamp amplifier (Axon Instruments) that also supplied the positive driving voltage. A lipid bilayer was formed across the aperture using 1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC) or 1,2-di-O-phytanyl-sn-glycero-3-phosphocholine (DOPC), obtained from Avanti Polar Lipids.

Following bilayer formation, M2-NNN-MspA was added to the cis chamber to a final concentration of ~2.5 ng ml⁻¹. A single pore insertion into the bilayer was recognized by a characteristic increase in the conductance. Upon single pore insertion, the cis chamber buffer was perfused out and replaced with MspA-free buffer to prevent the insertion of additional pores. The Hel308 motor enzyme was added to the cis chamber to a final concentration of ~50 nM, and DNA was added to a final concentration of ~5 nM.

Hel308 is used as a translocase, rather than a helicase, in the sequencing experiments presented here, similarly to experiments described previously36. In brief, Hel308 loads onto the overhanging 3′ end of the template DNA strand at the single-stranded–double-stranded junction. The 5′ end of the template strand is captured by the pore, and the blocking strand is sheared off as the template strand is pulled through the pore. Hel308 is too large to fit through MspA, and arrest the template strand translocation once the duplexed blocking strand has been completely sheared away. Hel308 proceeds as a translocase from 3′ to 5′ along the template strand, incrementally pulling the DNA out of the pore towards cis.

Operating buffers. All experiments were conducted using symmetric cis and trans buffer conditions of 400 mM KCl with 10 mM HEPES at pH 8.00 ± 0.05. The cis buffer also contained 1 mM EDTA, 1 mM DTT, 10 mM MgCl2, and 100 μM ATP. ATP-containing buffer was re-perfused into cis approximately once per hour to prevent depletion of ATP and accumulation of ADP. Experiments were performed at 20 °C.

Data acquisition and analysis. Experiments were controlled and data were acquired with custom acquisition software written in LabView (National Instruments, version 2018) at a sampling rate of 50 kHz. The ionic current signal was low-pass filtered at 10 kHz in the patch clamp amplifier. Ionic current traces were analyzed using custom programs written in Matlab (Mathworks, version 2018a).

Reads were filtered using a custom compression filter (Supplementary Note 14) to eliminate transient fluctuations in ionic current unrelated to translocating DNA sequence. Enzyme-controlled DNA translocation events were detected with a thresholding algorithm as described in previous work36. For constant-voltage experiments, the open-pore ionic current value was determined for the data, and an event was called whenever the ionic current returns to greater than 94% of the open-pore value. Events failing certain basic criteria (duration longer than 1 s, an average ionic current less than 10% or greater than 70% of the open-pore value) were automatically discarded. The remaining events were classified by eye to select events with a large number of enzyme steps. The same thresholding method was used for event detection in the variable-voltage data, with the sole difference being that the variable-voltage data were first downsampled to 200 Hz, thus removing the periodic characteristics of the signal.

Small variations in temperature, salt concentration, and electrode offsets from day to day, pore to pore, and read to read cause changes in both the overall magnitude of the observed conductances (an ‘offset’) as well as the relative magnitudes of adjacent states (a ‘scale’). We calibrate each read to the 6-mer model prior to sequencing using a scale and an offset calculated specifically for that read.

Statistics. A complete account of the number of reads collected on each DNA strand can be found in Supplementary Note 15.

In Fig. 1e, the uncertainty in the position shift as a function of voltage was determined using a bootstrapping method. The overall position shift was determined via analysis of the consensus signal of 18 variable-voltage reads of the same DNA sequence, as described in Supplementary Note 1. Using 10 unique subsets drawn from these 18 reads, we conducted identical analyses of the position shift as a function of voltage. The reported uncertainty (shaded region around the shift curve) is the standard deviation of these bootstrapped measurements.

In Fig. 2, the uncertainty around each conductance measurement (shaded regions) was determined as follows. For variable-voltage measurements (Fig. 2a,b, bottom panels, and Fig. 2c,d) we determined the covariance of the three principal component coefficients characterizing each segment by taking the covariance over the independent measurements of these three coefficients collected during each half-cycle of the voltage through the duration of that enzyme step (Supplementary Note 3). We converted the associated covariance of each mean conductance curve to a standard deviation around the mean by taking 100 random draws from a multivariate normal distribution with matching mean and covariance, then taking the standard deviation of these 100 random curves at each DNA position. For the constant-voltage measurements in the top panels of Fig. 2a,b, the shaded regions represent the standard deviation around the mean conductance extracted from the variable-voltage data at each enzyme step at the DNA position corresponding to the constant-voltage operating value of 180 mV.

For the determination of constant-voltage sequencing accuracy, we measured the average identity rate over all 9,368 bases sequenced with this method over 31 separate reads. The uncertainty in the overall accuracy was determined using a binomial error model as discussed in the main text. The average variable-voltage sequencing accuracy, as well as its associated uncertainty, was calculated in the same fashion, using all 17,309 bases sequenced over 97 separate reads. The confusion matrices in Fig. 3c,d broke down the sequencing results by base identity. For constant-voltage sequencing, the 9,368 total calls broke down into 2,203 As, 2,188 Cs, 2,166 Gs, 2,144 Ts, and 667 gaps. For variable-voltage sequencing, the 17,309 total calls broke down into 4,021 As, 4,073 Cs, 3,813 Gs, 4,081 Ts, and 1,321 gaps.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data for the main text figures as well as for all the constant- and variable-voltage sequencing reads used for the validation study can be found on figshare at https://doi.org/10.6084/m9.figshare.7723214. This also contains all of the Matlab code and supporting files that are necessary to replicate the sequencing analysis for both constant- and variable-voltage, as well as the scripts to generate the main text figures from their underlying data.

Code availability

Code and supporting files for constant- and variable-voltage sequencing analysis, as well as for main text figure generation can be found on github at https://github.com/uvwnanopore/variable-voltage-sequencing.git.

References

36. Laszlo, A. H., Derrington, I. M. & Gundlach, J. H. MspA nanopore as a single-molecule tool: from sequencing to SPRNT. *Methods* 105, 75–89 (2016).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- The exact sample size \((n)\) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. \(F, t, r\)) with confidence intervals, effect sizes, degrees of freedom and \(P\) value noted
- Given \(P\) values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's \(d\), Pearson's \(r\)), indicating how they were calculated

Clearly defined error bars

State explicitly what error bars represent (e.g. SD, SE, CI)

Software and code

Policy information about availability of computer code

| Data collection | Data was collected with custom data acquisition software written in LabView (National Instruments, version 2018). Custom software is available upon request. |
|----------------|------------------------------------------------------------------------------------------------------------------------------------------|
| Data analysis  | Data analysis was conducted using custom software written in Matlab (the Mathworks, version 2018a). Custom software is available upon request. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No statistical methods were used to pre-calculate sample sizes in the constant- and variable-voltage sequencing verification experiments. Sample sizes were chosen as the net yield in reads of several days of nanopore experiments. We terminated data collection on both experiments once we had measured read accuracy on multiple pores (N = 10 and N = 21 for constant and variable voltage, respectively) and over enough bases to accurately estimate the per-base error rates (N = 9368 bases and N = 17309 bases for constant and variable voltage, respectively). Likewise, no pre-calculation was used to determine sample sizes in the DNA stretching experiments. Enough samples were collected to generate a consensus signal for the target DNA sequence.

Data exclusions

In the sequencing verification experiment, reads failing certain basic quality requirements and reads of non-target DNA strands were excluded as follows. Reads exhibiting substantial long-duration mid-read gating (spontaneous drops in conductance not related to DNA sequence) were discarded prior to sequencing. Reads not containing a sufficient length of target DNA (>40 bp), either due to premature release by the motor enzyme or due to the capture of a small fragment from the restriction digest, were discarded from the final accuracy analysis. Adapter dimers, in which the enzyme-loading and pore-threading adapters used to facilitate capture and reading of the target DNA bind together without a payload of the target sequence in between (Supplemental Note 14) were discarded, as they did not contain the DNA sequence under study. We also observed and discarded a population of longer reads that did not have any significant alignment against the target pET28a reference sequence but did align to the E. coli reference genome. As the target pET28a DNA was grown up in an E. coli host, we concluded that these longer reads were reads of genomic E. coli DNA rather than the target sequence under study, and thus excluded them from our final accuracy analysis. The inclusion conditions of read length and lack of mid-event gating were established prior to data collection. The existence of non-pET28a reads due to the presence of E. coli DNA was not foreseen, and the exclusion of these reads was determined after data were collected.

Replication

Our sequencing verification experiments used data from multiple days and multiple pores, and evaluated sequencing accuracy over a range of sequence contexts (as contained in the pET28a reference sequence). We found that both constant- and variable-voltage sequencing accuracies were stable across different days and pores. As sequence context is known to affect nanopore sequencing accuracy, both methods were tested on the same target DNA sequence to provide a fair comparison of their relative performance. Different DNA sequence contexts yielding a higher (lower) accuracy for one of the two methods would similarly yield a higher (lower) accuracy for the other method. As this work is concerned with the relative performance improvement between the two methods, and the performance improvement was demonstrated to be stable across days, pores, and the presented sequence contexts, we conclude that a similar improvement would be observed on different target DNA sequences as well.

Randomization

Samples were not allocated into experimental groups as randomization was not relevant to our study.

Blinding

Investigators were not blinded as to the DNA sequence during the sequence verification experiments. Knowledge of the DNA sequence under study was necessary to the design of the experiment and preparation of the DNA for nanopore sequencing (choice of restriction enzymes and sequence of adapter strands; Supplemental Note 14). The data analysis and basecalling processes were put in place with fixed parameters prior to knowledge of the DNA sequence for the final verification experiment. Parameters were not tuned to generate best results on the target DNA.

Reporting for specific materials, systems and methods
### Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☒  | Unique biological materials |
| ☒  | Antibodies |
| ☒  | Eukaryotic cell lines |
| ☒  | Palaeontology |
| ☒  | Animals and other organisms |
| ☒  | Human research participants |

### Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☒  | ChIP-seq |
| ☒  | Flow cytometry |
| ☒  | MRI-based neuroimaging |