

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☒ ☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☒ ☐ A statement on 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
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ 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 statistical parameters 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
Give 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

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

No software were used for data collection.

Data analysis

The manuscript describes the hifiasm (ONT) assembly algorithm, which is freely available at: <https://github.com/chhylp123/hifiasm>. The specific version used in this study has been archived in a DOI-minted repository at <https://doi.org/10.5281/zenodo.18079611>. Assemblies using ONT Simplex and PacBio HiFi reads were generated with hifiasm (version 0.25.0-r726). For comparison, assemblies from ONT Simplex reads were also generated using HERRO (R10.4.1 model, <https://github.com/lbcb-sci/herro>) and Verkko (version v2.2.1). Seqtk (version 1.4-r130-dirty) was used to downsample HERRO-corrected reads to the target coverage for Verkko input. To identify telomere-to-telomere (T2T) contigs and scaffolds, we used a script provided by the Human Pangenome Reference Consortium (HPRC): <https://github.com/biomonika/HPP/blob/main/assembly/wdl/workflows/assessAssemblyCompleteness.wdl>. K-mer-based assembly quality was evaluated with yak (version 0.1-r62-dirty) and Merquy (version 1.3). Meryl (version 1.4.1) was used together with Merquy to build the k-mer index. For human genome assemblies, gene completeness was assessed by aligning cDNAs to both the reference genome and the assembled contigs using minimap2 (version 2.28-r1209). Gene-level completeness was then evaluated using paftools.js, a utility included in the minimap2 package. The alignments used for SVbyEye visualization (commit fdee406) were also generated with minimap2 (version 2.28-r1209). Immuannot (commit 31362c3) was used to annotate immunological genes. Assembly-based variant calling was performed using Dippyall v0.3 and benchmarked against the GIAB HG002 T2TQ100-v1.0 truth set and its corresponding confident regions on GRCh38. Benchmarking of the resulting variant calls was then carried out using hap.py v0.3.15. VCF files were processed using bcftools (version 1.21) and tabix from SAMtools (version 1.22.1). Flagger v1.1.0 and NucFlag v0.3.6 were used to detect misassemblies in a reference-free manner. To improve alignment accuracy in highly repetitive regions—particularly for reference-based disassembly detection—we employed a hybrid strategy combining MashMap v3.1.3 and Minimap2 v2.30-r1287. Nucleotide composition at each contig end was calculated using bedtools nuc v2.31.0. bedtools was also used to identify assembly gaps. IGV v2.16.0 was used to visualize read and assembly alignments.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Human reference genome: GRCh38, CHM13v2, and HG002 Q100; ONT standard reads of HG001 (SUP): [s3://ont-open-data/giab_2025.01/basecalling/sup/HG001](https://ont-open-data/giab_2025.01/basecalling/sup/HG001); ONT standard reads of HG001 (HAC): [s3://ont-open-data/giab_2025.01/basecalling/hac/HG001](https://ont-open-data/giab_2025.01/basecalling/hac/HG001); PacBio HiFi reads of HG001: https://ftp.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/NA12878/HudsonAlpha_PacBio_CCS/; Illumina short reads of HG001 and parents: ERR194147 (HG001), ERR194160 (paternal), ERR194161 (maternal); ONT standard reads of HG002 (SUP): [s3://ont-open-data/giab_2025.01/basecalling/sup/HG002/PAW70337](https://ont-open-data/giab_2025.01/basecalling/sup/HG002/PAW70337); ONT standard reads of HG002 (HAC): [s3://ont-open-data/giab_2025.01/basecalling/hac/HG002/PAW70337](https://ont-open-data/giab_2025.01/basecalling/hac/HG002/PAW70337); ONT ultra-long reads of HG002: https://s3-us-west-2.amazonaws.com/human-pangenomics/index.html?prefix=submissions/5b73fa0e-658a-4248-b2b8-cd16155bc157--UCSC_GIAB_R1041_nanopore/HG002_R1041_UL/dorado/v0.4.0_wMods/*ULCIR*.bam and https://s3-us-west-2.amazonaws.com/human-pangenomics/index.html?prefix=submissions/5b73fa0e-658a-4248-b2b8-cd16155bc157--UCSC_GIAB_R1041_nanopore/HG002_R1041_UL/dorado/v0.4.0_wMods/*ULNEB*.bam; PacBio HiFi reads of HG002: https://ftp.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/HG002_NA24385_son/PacBio_HiFi-Revio_20231031/; Illumina short reads of HG002 and parents: https://s3-us-west-2.amazonaws.com/human-pangenomics/index.html?prefix=working/HPRC_PLUS/HG002/raw_data/Illumina/; ONT standard reads of HG003 (SUP): [s3://ont-open-data/giab_2025.01/basecalling/sup/HG003](https://ont-open-data/giab_2025.01/basecalling/sup/HG003); ONT standard reads of HG003 (HAC): [s3://ont-open-data/giab_2025.01/basecalling/hac/HG003](https://ont-open-data/giab_2025.01/basecalling/hac/HG003); PacBio HiFi reads of HG003: https://ftp.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/HG003_NA24149_father/PacBio_HiFi-Revio_20231031/ and https://ftp.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/HG003_NA24149_father/PacBio_CCS_Google_15kb/; Illumina short reads of HG003: https://s3-us-west-2.amazonaws.com/human-pangenomics/index.html?prefix=working/HPRC_PLUS/HG002/raw_data/Illumina/parents/HG003/; ONT standard reads of HG004 (SUP): [s3://ont-open-data/giab_2025.01/basecalling/sup/HG004](https://ont-open-data/giab_2025.01/basecalling/sup/HG004); ONT standard reads of HG004 (HAC): [s3://ont-open-data/giab_2025.01/basecalling/hac/HG004](https://ont-open-data/giab_2025.01/basecalling/hac/HG004); PacBio HiFi reads of HG004: https://ftp.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/HG004_NA24143_mother/PacBio_HiFi-Revio_20231031/ and https://ftp.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/HG004_NA24143_mother/PacBio_CCS_Google_15kb/; Illumina short reads of HG004: https://s3-us-west-2.amazonaws.com/human-pangenomics/index.html?prefix=working/HPRC_PLUS/HG002/raw_data/Illumina/parents/HG004/; ONT standard reads of HG005 (SUP): [s3://ont-open-data/giab_2025.01/basecalling/sup/HG005](https://ont-open-data/giab_2025.01/basecalling/sup/HG005); ONT standard reads of HG005 (HAC): [s3://ont-open-data/giab_2025.01/basecalling/hac/HG005](https://ont-open-data/giab_2025.01/basecalling/hac/HG005); PacBio HiFi reads of HG005: https://ftp.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/ChineseTrio/HG005_NA24631_son/PacBio_CCS_15kb_20kb_chemistry2/uBAMs/; Illumina short reads of HG005: https://s3-us-west-2.amazonaws.com/human-pangenomics/index.html?prefix=working/HPRC_PLUS/HG005/raw_data/Illumina/; ONT standard reads of HG006 (SUP): [s3://ont-open-data/giab_2025.01/basecalling/sup/HG006](https://ont-open-data/giab_2025.01/basecalling/sup/HG006); ONT standard reads of HG006 (HAC): [s3://ont-open-data/giab_2025.01/basecalling/hac/HG006](https://ont-open-data/giab_2025.01/basecalling/hac/HG006); PacBio HiFi reads of HG006: https://ftp.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/ChineseTrio/HG006_NA24694-huCA017E_father/PacBio_CCS_15kb_20kb_chemistry2/uBAMs/ and https://ftp.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/ChineseTrio/HG006_NA24694-huCA017E_father/PacBio_HiFi_Google/; Illumina short reads of HG006: https://s3-us-west-2.amazonaws.com/human-pangenomics/index.html?prefix=working/HPRC_PLUS/HG005/raw_data/Illumina/parents/HG006/; ONT standard reads of HG007 (SUP): [s3://ont-open-data/giab_2025.01/basecalling/sup/HG007](https://ont-open-data/giab_2025.01/basecalling/sup/HG007); ONT standard reads of HG007 (HAC): [s3://ont-open-data/giab_2025.01/basecalling/hac/HG007](https://ont-open-data/giab_2025.01/basecalling/hac/HG007); PacBio HiFi reads of HG007: https://ftp.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/ChineseTrio/HG007_NA24695-hu38168_mother/PacBio_CCS_15kb_20kb_chemistry2/uBAMs/ and https://ftp.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/ChineseTrio/HG007_NA24695-hu38168_mother/PacBio_HiFi_Google/; Illumina short reads of HG007: https://s3-us-west-2.amazonaws.com/human-pangenomics/index.html?prefix=working/HPRC_PLUS/HG005/raw_data/Illumina/parents/HG007/; ONT ultra-long reads of HG02818: https://s3-us-west-2.amazonaws.com/human-pangenomics/index.html?prefix=working/HPRC_PLUS/HG02818/raw_data/Illumina/; ONT ultra-long reads of Arabidopsis thaliana: SRR29061597; Illumina short reads of Arabidopsis thaliana: <https://ngdc.cncb.ac.cn/gsa/browse/CRA005350>; ONT ultra-long data of Danio rerio: https://genomeark.s3.amazonaws.com/index.html?prefix=species/Danio_rerio/fDanRer17/genomic_data/ont/pod5/ and https://genomeark.s3.amazonaws.com/index.html?prefix=species/Danio_rerio/fDanRer17/genomic_data/ont/fast5/; Illumina short reads of Danio rerio: https://www.ncbi.nlm.nih.gov/sra?linkname=bioproject_sra_all&from_uid=1029986; ONT ultra-long data of Solanum lycopersicum: https://obj.umiacs.umd.edu/marbl/publications/duplex/Solanum_lycopersicum_heinz1706/UL/R10.4_40x.noduplex.fastq.gz; Illumina short reads of Solanum lycopersicum: <https://ngdc.cncb.ac.cn/gsa/browse/CRA003995/CRX232533>; ONT reads of Linum usitatissimum (Flax): SRR31124331; ONT reads used for comparison with Napu (Shasta): https://s3-us-west-2.amazonaws.com/human-pangenomics/index.html?prefix=publications/Napu_paper_ONT_Coriell_SingleFC_2023/HG0*_R10/reads/*_bam; Napu (Shasta) assemblies:

genome assemblies from ONT reads), <https://zenodo.org/records/17613526> (non-human genome assemblies from ONT reads), and <https://zenodo.org/records/15205178> (assemblies from PacBio HiFi reads). All links and accession identifiers for the sequencing data are also listed in Supplementary Table 5.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Reporting on race, ethnicity, or other socially relevant groupings

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

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

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

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

Sample size

Data exclusions

Replication

Randomization

Blinding

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Plants

Seed stocks	Not Applicable
Novel plant genotypes	Not Applicable
Authentication	Not Applicable