PacBio

Iso-Seq

Scalable
De Novo
Isoform Discovery
from PacBio HiFi Reads

Oxford NANOPORE™ Technologies

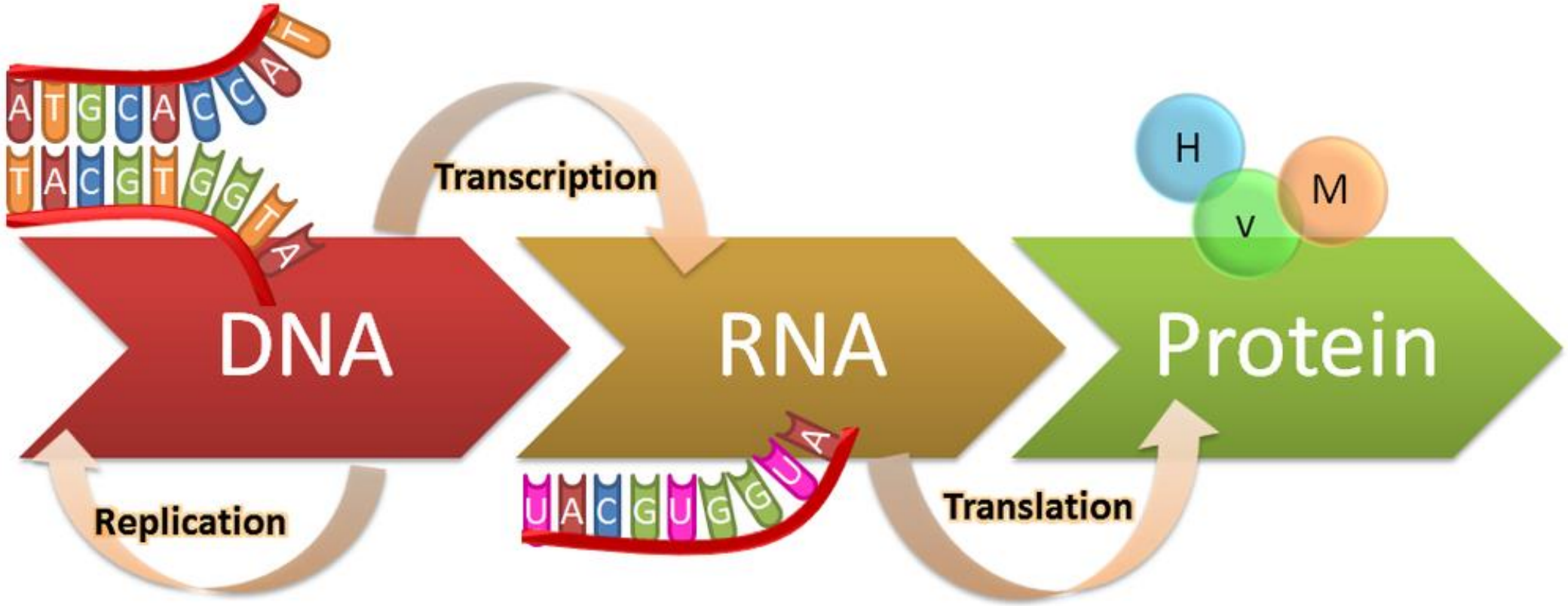
AN OVERVIEW OF LONG-READ SEQUENCING

Amanda Markee | November 14th 2024 | CG2

OUTLINE

- INTRODUCTION
- HISTORY OF LONG-READ SEQUENCING
- PROS AND CONS
- EXAMPLES OF RECENT LITERATURE
- CONCLUSIONS

INTRODUCTION



DNA = GENETIC BLUEPRINT

PROTEIN = PRODUCT

INTRODUCTION

DNA = GENETIC BLUEPRINT

PROTEIN = PRODUCT

INTRODUCTION

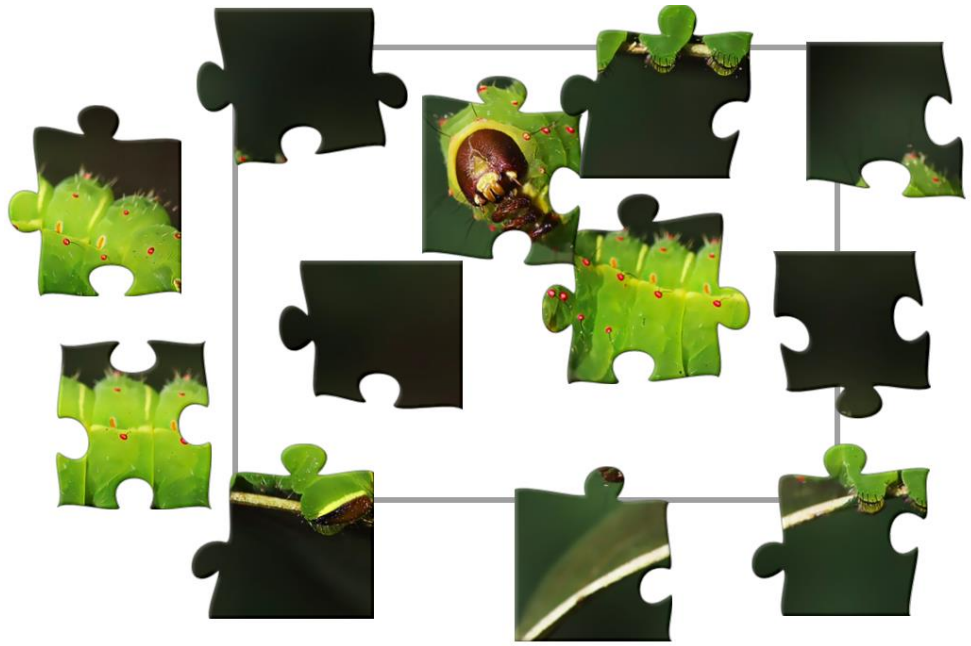
Short-read sequencing (NGS) produces many short DNA fragments (150bp – 300bp)

Long-read sequencing produces longer, contiguous DNA fragments (5,000bp – 30,000 bp)

INTRODUCTION

Short-read sequencing (NGS) produces many short DNA fragments (150bp – 300bp)

Long-read sequencing produces longer, contiguous DNA fragments (5,000bp – 30,000 bp)



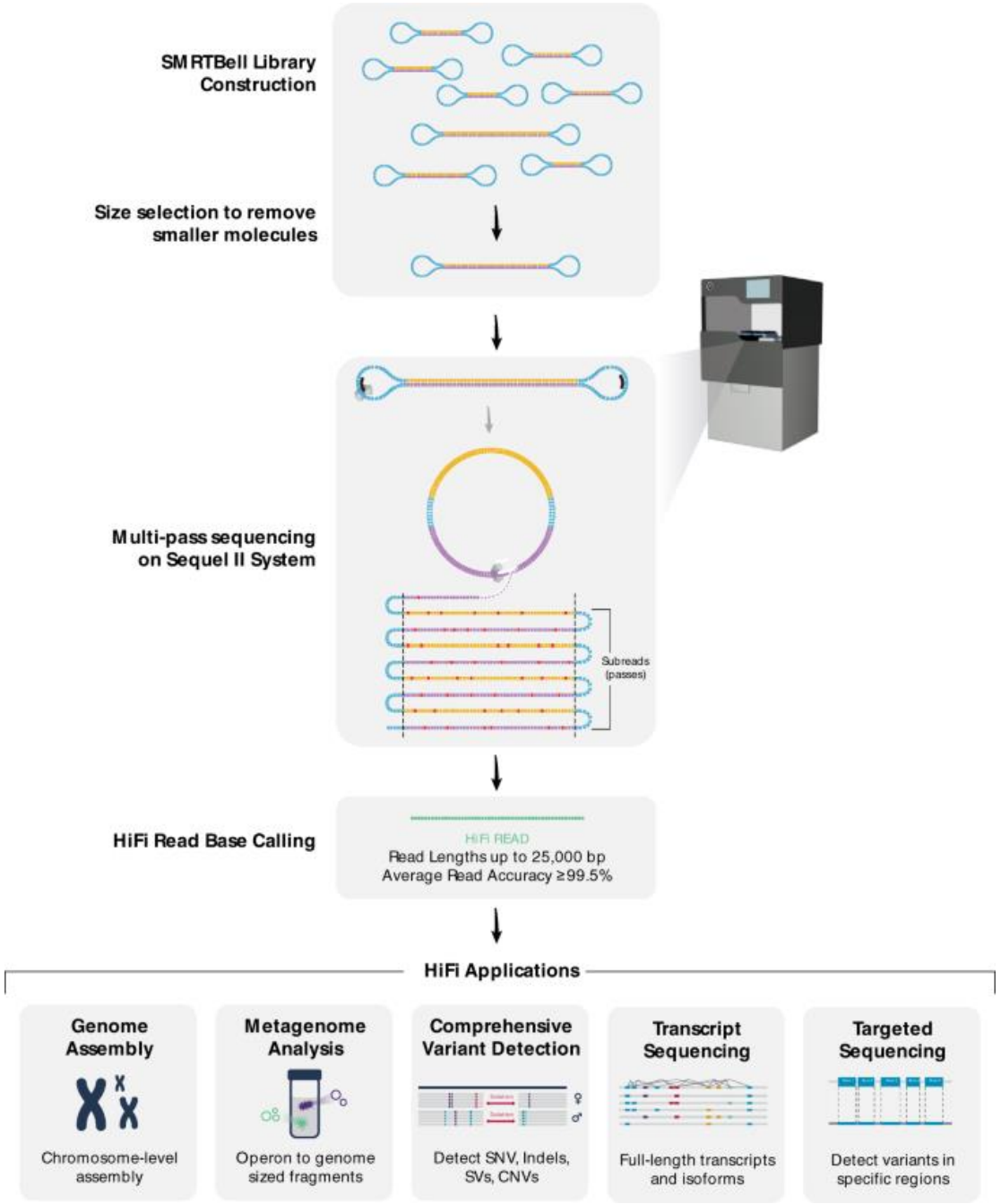
INTRODUCTION

Long-read sequencing for de-novo genome assembly, capturing complex genomic regions (structural variants, isoforms, etc.), or highly repetitive regions.

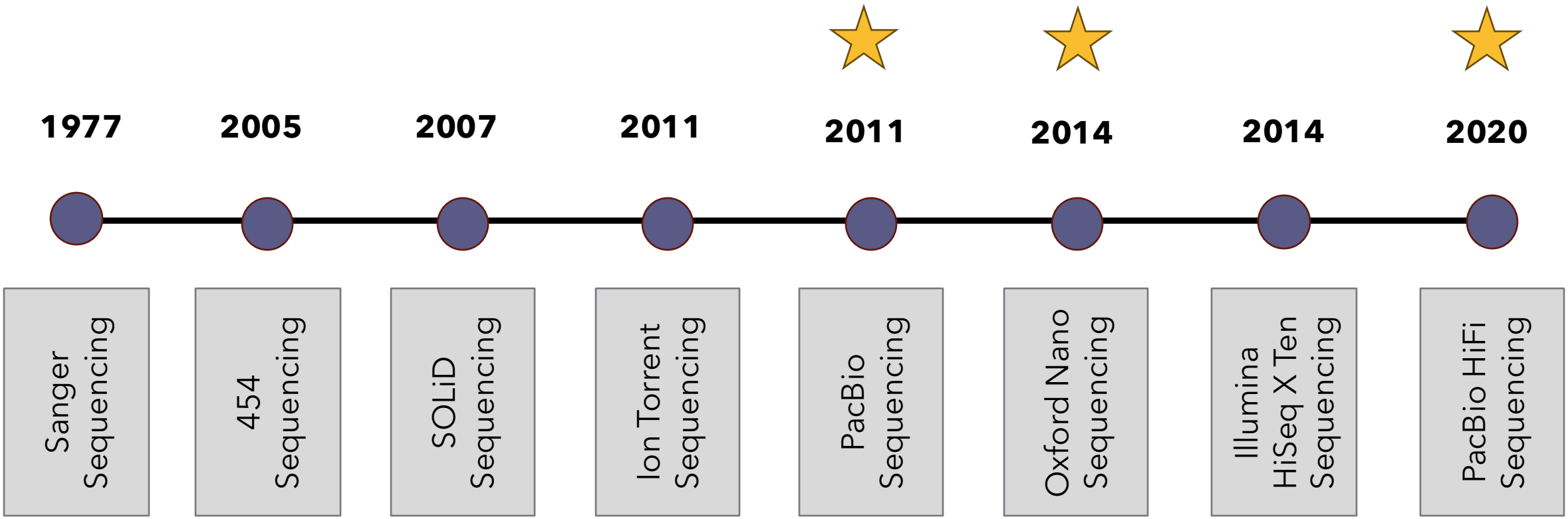
SMRT cell (PacBio) sequencing takes place on a chip lined with individual wells

DNA polymerase is used to sequence a complementary strand, and the fluorescence is measured to identify the corresponding nucleotides.

CCS reads (PacBio) ensure high accuracy



HISTORY OF LONG-READ SEQUENCING



PROS

•**Accuracy and Resolution**: CCS reads allow for equally as accurate reads as NGS with high accuracy up to 99.5%
•**Improved Assembly**: Long reads allow for a more complete and contiguous assembly, less gaps than short reads, especially useful for *de-novo* genome assembly.
•**Applications in Challenging Genomes**: Useful in sequencing large repetitive genomes such as plant or squamate genomes. Easily resolves repetitive regions

CONS

•**Cost**: If going for high-quality, can be more costly (1 SMRT Cell ~ $3000; 1 NGS lane ~ $800)
•**Error Rate**: Long read technologies tend to have higher raw error rates than short-read sequencing, though error correction methods can mitigate this (CCS)
•**Technical Limitations**: sample preparation more specialized (HMW DNA, needs little degradation, etc. harder for in-field work, or degraded DNA samples like museum specimens)

Rapid molecular diversification and homogenization of clustered major ampullate silk genes in Argiope garden spiders

Richard H Baker [1], André Corvelo [2], Cheryl Y Hayashi [1]

Affiliations + expand

PMID: 36508456 PMCID: PMC9779670 DOI: 10.1371/journal.pgen.1010537

Long Reads Are Revolutionizing 20 Years of Insect Genome Sequencing

Scott Hotaling [1,*], John S. Sproul[2], Jacqueline Heckenhauer[3,4], Ashlyn Powell[5], Amanda M. Larracuente[2], Steffen U. Pauls[3,4,6], Joanna L. Kelley [1], and Paul B. Frandsen [3,5,7,*]

[1]School of Biological Sciences, Washington State University, Pullman, Washington, USA
[2]Department of Biology, University of Rochester, New York, USA
[3]LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt, Germany
[4]Department of Terrestrial Zoology, Entomology III, Senckenberg Research Institute and Natural History Museum Frankfurt, Frankfurt, Germany
[5]Department of Plant and Wildlife Sciences, Brigham Young University, Provo, Utah, USA
[6]Institute for Insect Biotechnology, Justus-Liebig-University, Giessen, Germany
[7]Data Science Lab, Smithsonian Institution, Washington, District of Columbia, USA

*Corresponding authors: E-mails: scott.hotaling@wsu.edu; paul_frandsen@byu.edu.

Accepted: 10 June 2021

Evolution of Opsin Genes in Caddisflies (Insecta: Trichoptera)

Ashlyn Powell [1,*], Jacqueline Heckenhauer [2,3], Steffen U. Pauls [2,3], Blanca Ríos-Touma [4], Ryoichi B. Kuranishi [5,6], Ralph W. Holzenthal [7], Ernesto Razuri-Gonzales [3], Seth Bybee [8], Paul B. Frandsen [1,*]

Review Article | Published: 05 June 2020

Long-read human genome sequencing and its applications

Glennis A. Logsdon, Mitchell R. Vollger & Evan E. Eichler

Nature Reviews Genetics **21**, 597–614 (2020) | Cite this article

Darwin TREE of LIFE

Sequencing the genomes of 70,000 plants, fungi, animals and protists in Britain and Ireland

WHO WE ARE GENOMES DISCOVER GET INVOLVED FAQS

Darwin Tree of Life

The Darwin Tree of Life project aims to sequence the genomes of 70,000 species of eukaryotic organisms in Britain and Ireland.

It is a collaboration between biodiversity, genomics and analysis partners that is transforming the way we do biology, conservation and biotechnology.

EXAMPLE PAPER

JOURNAL ARTICLE

De Novo Long-Read Genome Assembly and Annotation of the Luna Moth (*Actias luna*) Fully Resolves Repeat-Rich Silk Genes

Amanda Markee ✉, Rebekah Keating Godfrey, Paul B Frandsen, Yi-Ming Weng, Deborah A Triant, Akito Y Kawahara

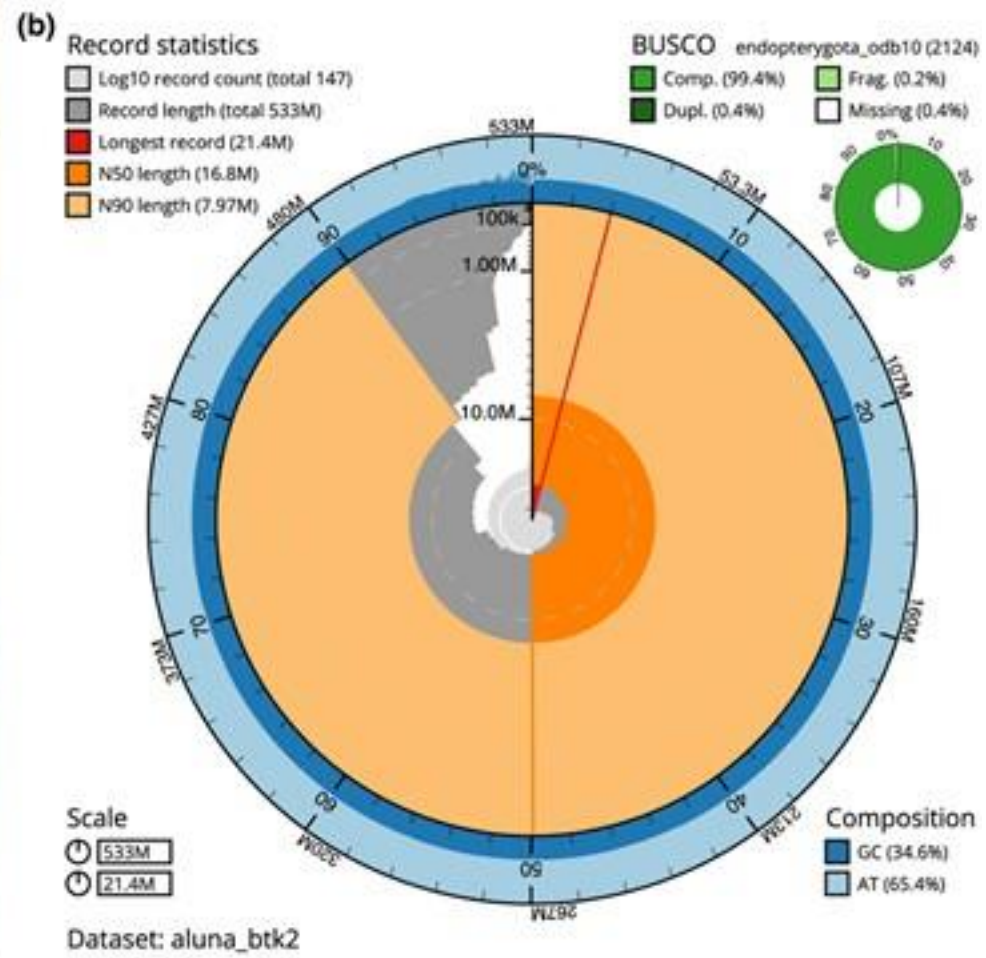
Genome Biology and Evolution, Volume 16, Issue 7, July 2024, evae148,
https://doi.org/10.1093/gbe/evae148
Published: 03 July 2024 **Article history** ▾

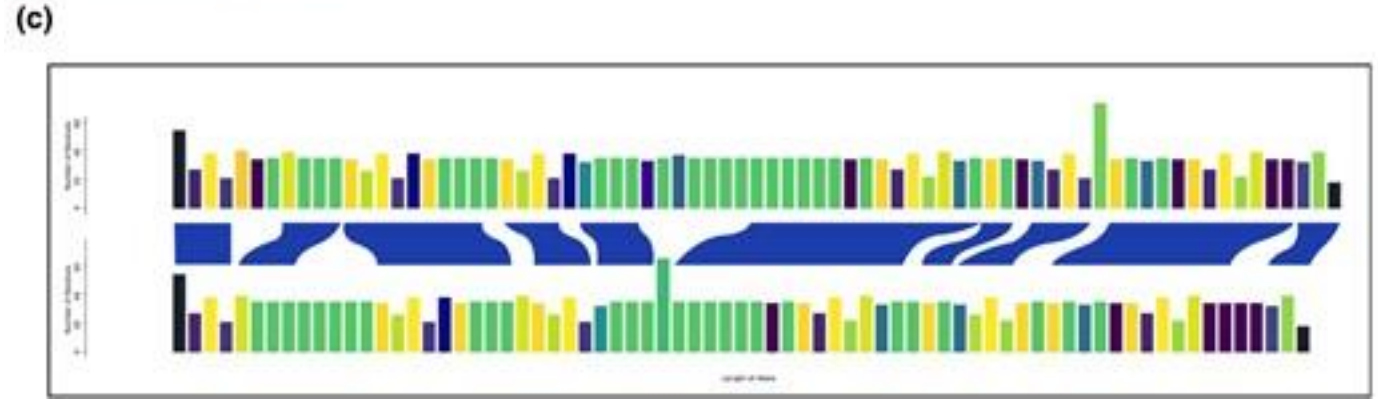
Table 2 A comparison of genome completeness between the only other existing *A. luna* genome assembly available to date

Open in new tab

	BUSCO complete (%)	Single Copy (%)	Duplicated (%)	Fragmented (%)	Missing (%)	Ref.
Actias luna[a]	99.4	99.0	0.4	0.2	0.4	Authors
Actias luna	71.4	63.9	7.5	15.4	13.2	GCA_010014465.3

[a]Genomes produced with long-read sequencing platforms, e.g. PacBio or Oxford Nanopore.



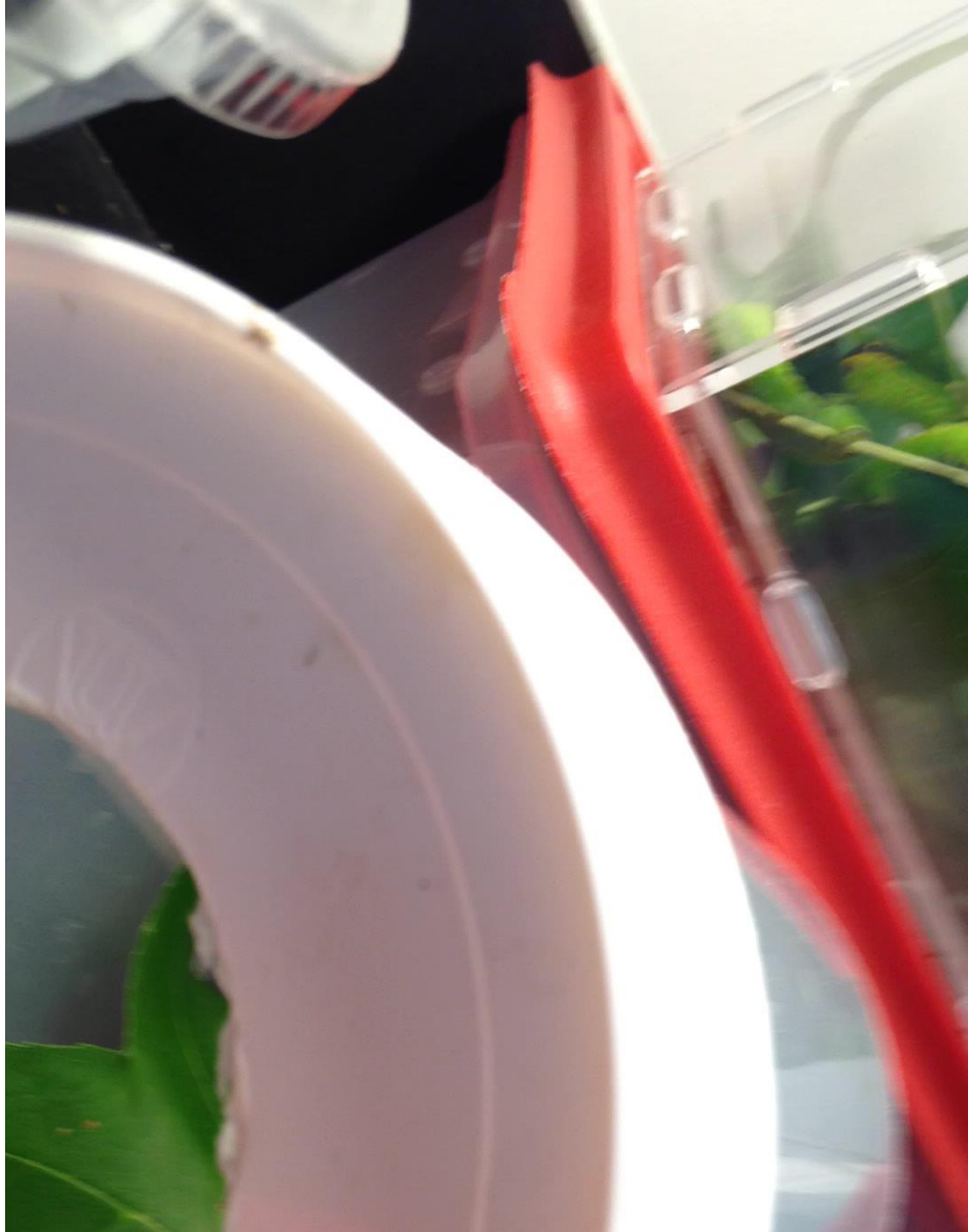


ASSEMBLY STATS EXAMPLE

Parameter	*Actias luna* (long-read)	*Actias luna* (short-read)
Reference	This study	2016 genome
Platform	PacBio Sequel IIe	Illumina MiSeq/HiSeq
Genome completeness	99.4%	71.4%
Number of contigs	155	541,894
Contig N50	16,802,800	2,189
Contig L50	14	64,346
GC content	34.65%	35.50%
Shortest contig (bp)	6,149	1
Longest contig (bp)	21,422,706	136,946
Mean contig (bp)	3,441,632	1055

CONCLUSIONS

- **Long read sequencing** has provided the opportunity for an influx of high-quality genomic resources for both model and non-model systems
- **Previous concerns** of lower accuracy continue to be addressed and built upon, making long-read sequencing arguably the most high-quality option for new genomes
- **Comparatively** long reads are **costly** and hopefully in time, will become more affordable and accessible across disciplines.



REFERENCES

Baker, R. H., Corvelo, A., & Hayashi, C. Y. (2022). Rapid molecular diversification and homogenization of clustered major ampullate silk genes in Argiope garden spiders. *PLOS Genetics*, *18*(12), e1010537. https://doi.org/10.1371/journal.pgen.1010537

Correa-Garhwal, S. M., Baker, R. H., Clarke, T. H., Ayoub, N. A., & Hayashi, C. Y. (2022). The evolutionary history of cribellate orb-weaver capture thread spidroins. *BMC Ecology and Evolution*, *22*(1), 89. https://doi.org/10.1186/s12862-022-02042-5

Ellis, E. A., Storer, C. G., & Kawahara, A. Y. (2021). *De novo* genome assemblies of butterflies. *GigaScience*, *10*(6), giab041. https://doi.org/10.1093/gigascience/giab041

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9451–9457. https://doi.org/10.1073/pnas.1921046117

Godfrey, R. K., Britton, S. E., Mishra, S., Goldberg, J. K., & Kawahara, A. Y. (2023). A high-quality, long-read genome assembly of the whitelined sphinx moth (Lepidoptera: Sphingidae: Hyles lineata) shows highly conserved melanin synthesis pathway genes. *G3 Genes|Genomes|Genetics*, *13*(6), jkad090. https://doi.org/10.1093/g3journal/jkad090

Heckenhauer, J., Stewart, R. J., Ríos-Touma, B., Powell, A., Dorji, T., Frandsen, P. B., & Pauls, S. U. (2023). Characterization of the primary structure of the major silk gene, h-fibroin, across caddisfly (Trichoptera) suborders. *iScience*, *26*(8), 107253. https://doi.org/10.1016/j.isci.2023.107253

Hotaling, S., Sproul, J. S., Heckenhauer, J., Powell, A., Larracuente, A. M., Pauls, S. U., Kelley, J. L., & Frandsen, P. B. (2021). Long Reads Are Revolutionizing 20 Years of Insect Genome Sequencing. *Genome Biology and Evolution*, *13*(8), evab138. https://doi.org/10.1093/gbe/evab138

Kawahara, A. Y., Storer, C. G., Markee, A., Heckenhauer, J., Powell, A., Plotkin, D., Hotaling, S., Cleland, T. P., Dikow, R. B., Dikow, T., Kuranishi, R. B., Messcher, R., Pauls, S. U., Stewart, R. J., Tojo, K., Frandsen, P. B., Storer, C. G., Markee, A., Heckenhauer, J., … Frandsen, P. B. (2022). Long-read HiFi sequencing correctly assembles repetitive heavy fibroin silk genes in new moth and caddisfly genomes. *Gigabyte*, *2022*, 1–14. https://doi.org/10.46471/gigabyte.64

Long-read human genome sequencing and its applications | Nature Reviews Genetics. (n.d.). Retrieved November 14, 2024, from https://www.nature.com/articles/s41576-020-0236-x

Markee, A., Godfrey, R. K., Frandsen, P. B., Weng, Y.-M., Triant, D. A., & Kawahara, A. Y. (2024). De Novo Long-Read Genome Assembly and Annotation of the Luna Moth (*Actias luna*) Fully Resolves Repeat-Rich Silk Genes. *Genome Biology and Evolution*, *16*(7), evae148. https://doi.org/10.1093/gbe/evae148

Marx, V. (2023). Method of the year: Long-read sequencing. *Nature Methods*, *20*(1), 6–11. https://doi.org/10.1038/s41592-022-01730-w

Powell, A., Heckenhauer, J., Pauls, S. U., Ríos-Touma, B., Kuranishi, R. B., Holzenthal, R. W., Razuri-Gonzales, E., Bybee, S., & Frandsen, P. B. (2024). Evolution of Opsin Genes in Caddisflies (Insecta: Trichoptera). *Genome Biology and Evolution*, *16*(9), evae185. https://doi.org/10.1093/gbe/evae185