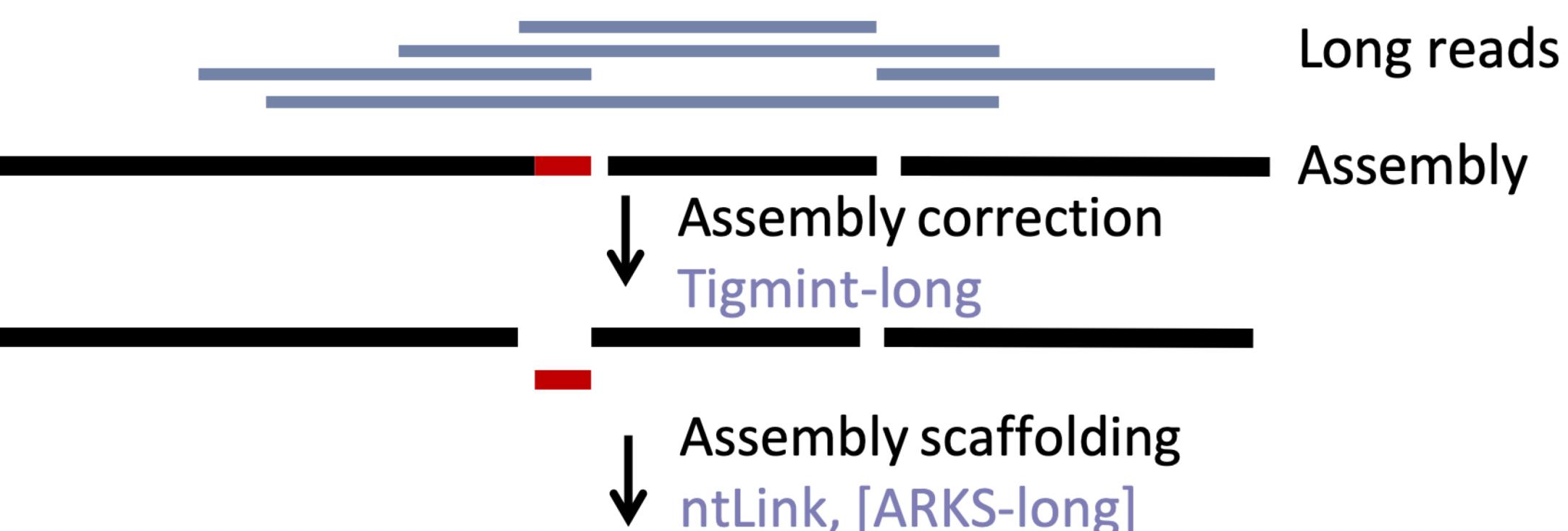


longStitch: High-quality genome assembly correction and scaffolding using long reads

Lauren Coombe, Janet X Li, Theodora Lo, Johnathan Wong, Vladimir Nikolic, René L Warren and Inanc Birol

lcoombe@bcgsc.ca
www.birollab.ca

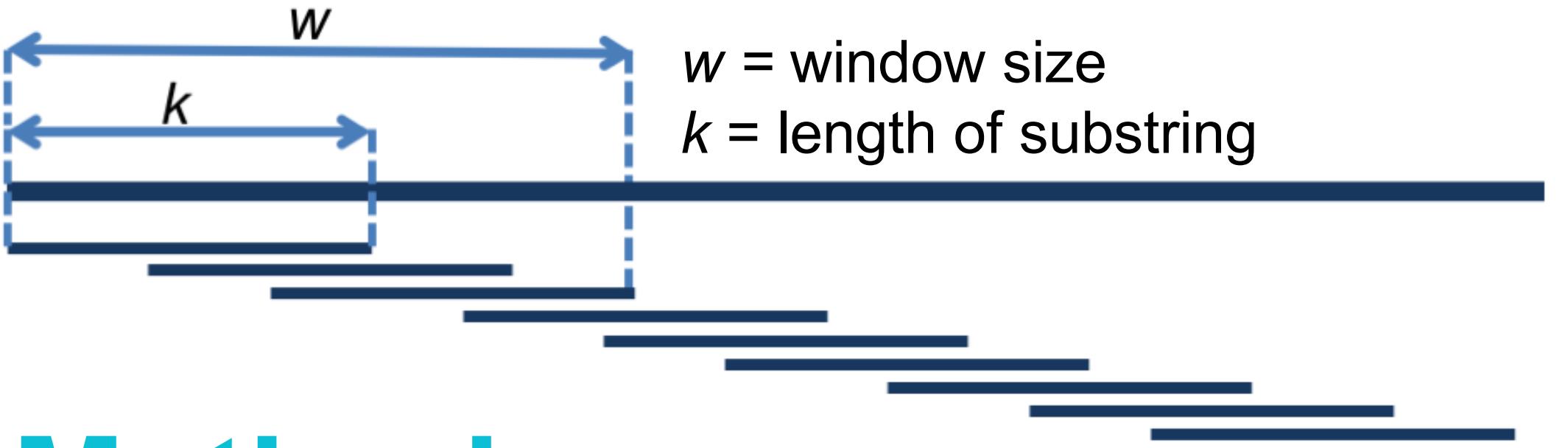
LongStitch



- Long read *de novo* assembly correction/scaffolding (3 steps)
- ntLink: newly developed long-read scaffolder
- Tigmint-long: correction
- ARKS-long: scaffolding

Minimizer sketches

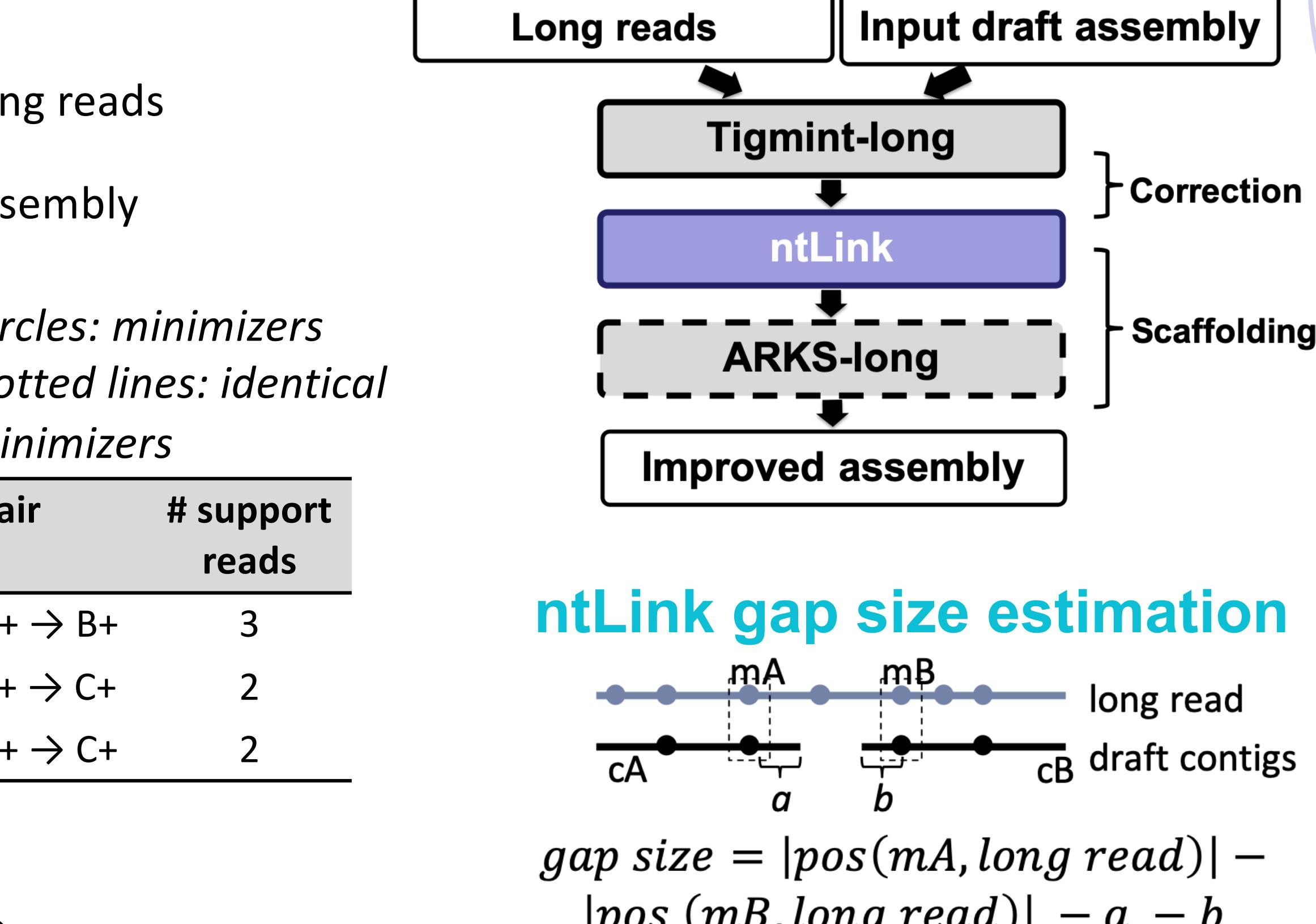
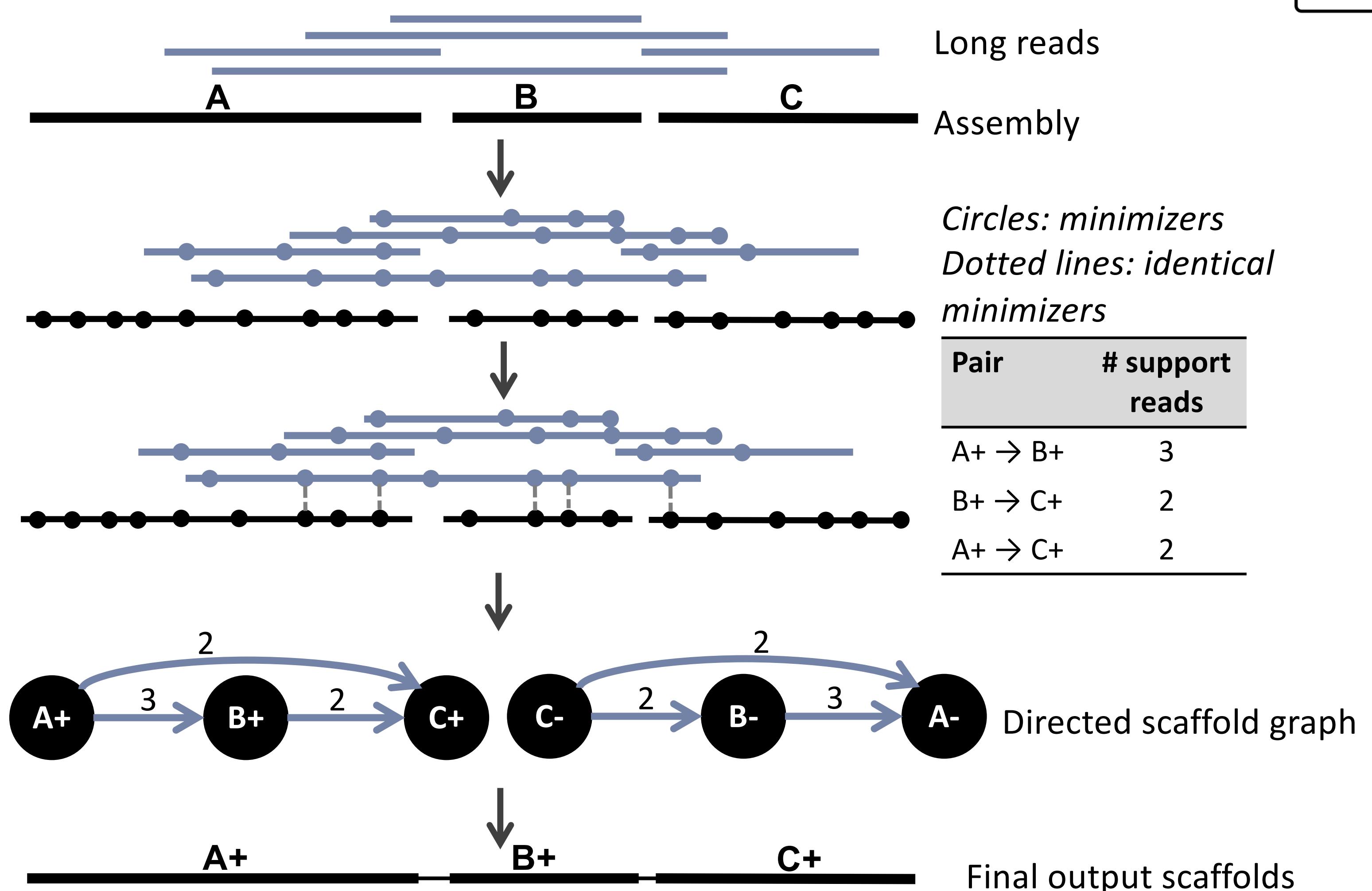
Reduce computational cost of sequence data storage and manipulation¹



- For each window of w adjacent k -mers:
- Compute hash values of each k -mer
 - Window's minimizer = smallest hash value
- Generates ordered list of minimizers per sequence

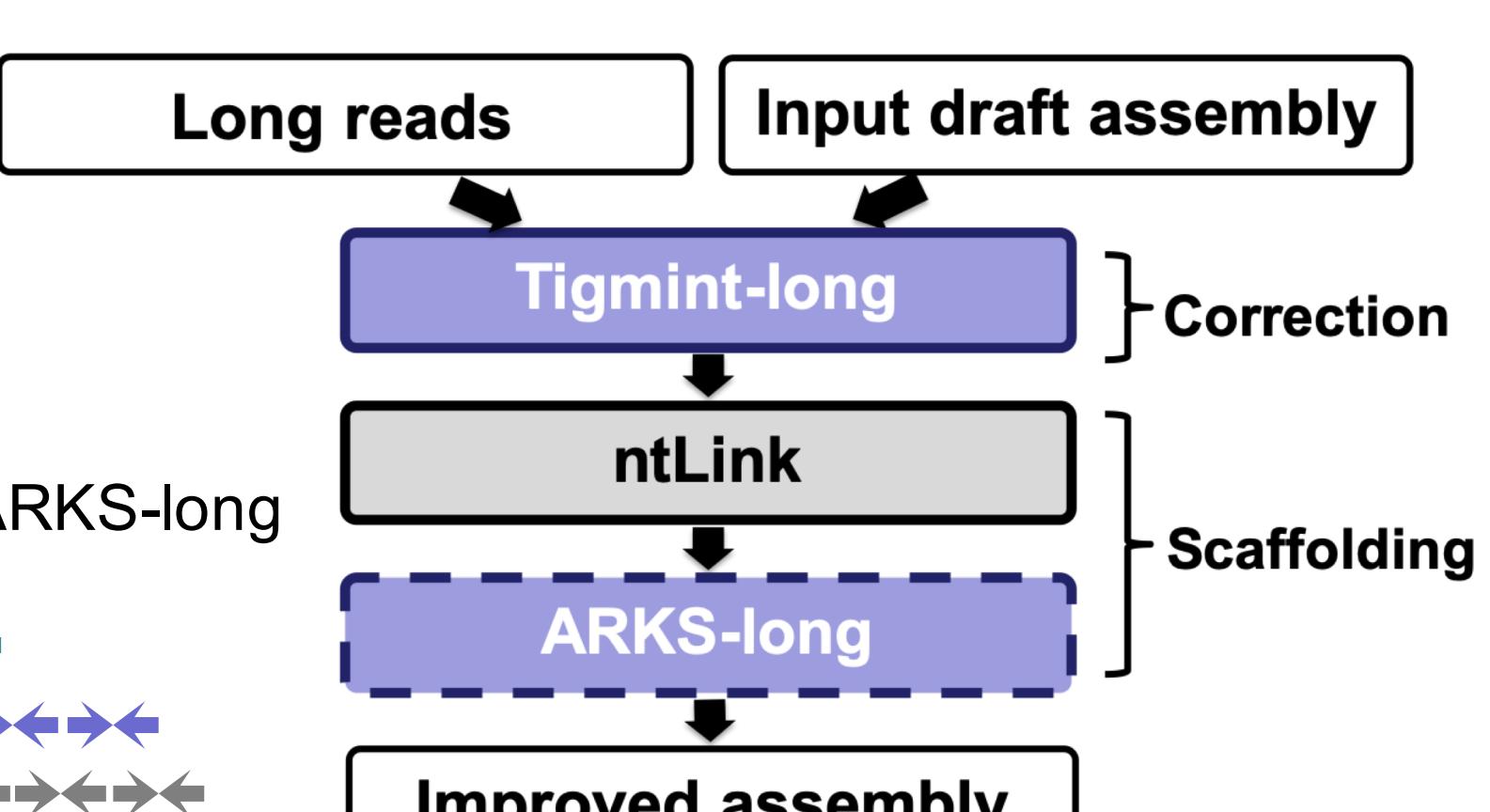
Methods

ntLink algorithm



Adapting Tigmint and ARKS for long reads

Originally developed for linked reads^{2,3,4}



Results

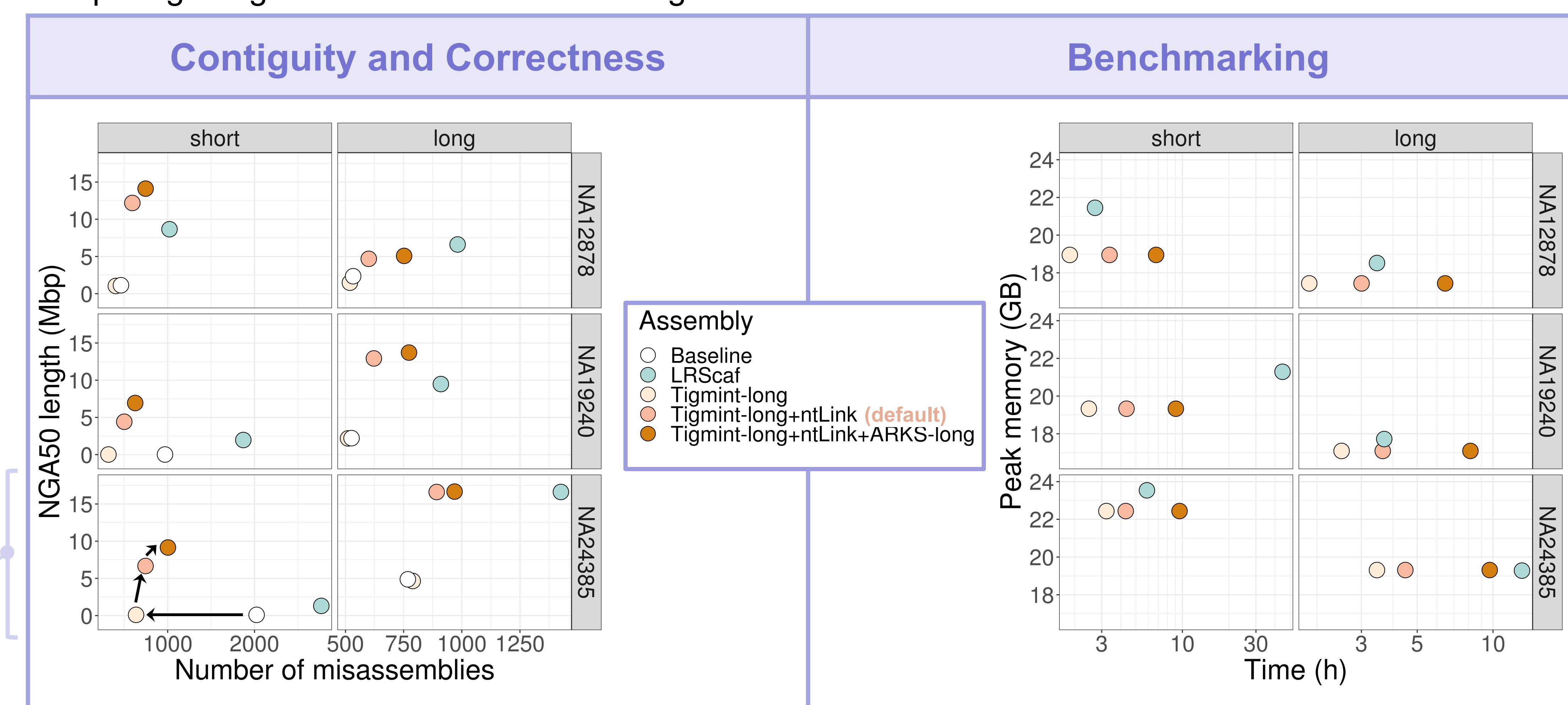
Running LongStitch (human data)

Short: short-read ABYSS⁵ assembly

Long: long-read Shasta⁶ assembly

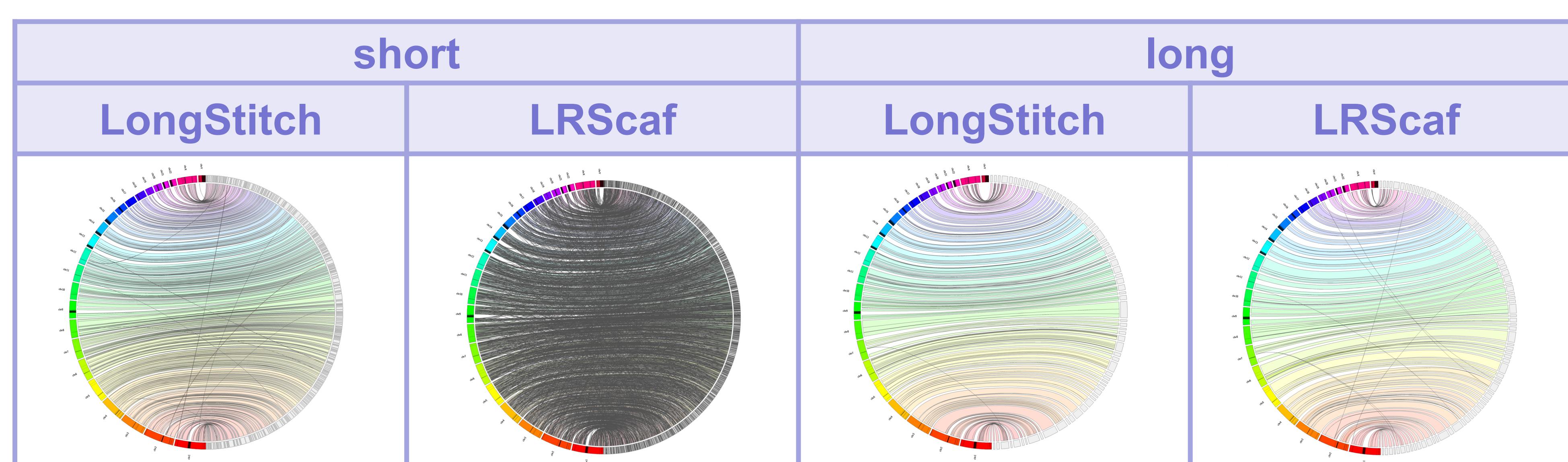
Comparing LongStitch to state-of-the-art long read scaffolder LRScf⁷

Individual	NA12878	NA19240	NA24385
Nanopore read coverage	39X	49X	51X



Visualizing large-scale misassemblies with Jupiter⁸ plots – NA24385

Large misassemblies evident as interrupting ribbons



Conclusions

- LongStitch: scalable assembly correction and scaffolding
- Leverages the rich information in long reads
- Generates high-quality genome assemblies
- Paper describing LongStitch: Coombe, L. et al. (2021) bioRxiv, 2021.06.17.448848.



Software Availability

<https://github.com/bcgsc/longstitch>

<https://github.com/bcgsc/ntlink>

`conda install -c bioconda longstitch`

Funding

Genome British Columbia

Genome Canada

National Institutes of Health

References

1. Roberts, M. et al. (2004) *Bioinformatics*, 20, 3363–3369.
2. Jackman, S.D. et al. (2018) *BMC Bioinformatics*, 19, 1–10.
3. Yeo, S. et al. (2018) *Bioinformatics*, 34, 725–731.
4. Coombe, L. et al. (2018) *BMC Bioinformatics*, 19, 1–10.
5. Jackman, S.D. et al. (2017) *Genome Res.*, 27, 768–777.
6. Shafin, K. et al. (2020) *Nat. Biotechnol.*, 38, 1044–1053.
7. Qin, M. et al. (2019) *BMC Genomics*, 20, 1–12.
8. Chu, J. (2018) *Zenodo*. doi:10.5281/ZENODO.1241235.