

RESEARCH NOTE

Open Access



# LAMPS: an analysis pipeline for sequence-specific ligation-mediated amplification reads

Christopher J. F. Cameron<sup>1,2\*</sup> , Xue Q. D. Wang<sup>2</sup>, Josée Dostie<sup>2</sup> and Mathieu Blanchette<sup>1</sup>

## Abstract

**Objective:** Ligation-Mediated Amplification (LMA) is a versatile biochemical tool for amplifying selected DNA sequences. LMA has increased in popularity due to its integration within chromosome conformation capture (5C) and chromatin immunoprecipitation (2C-ChIP) methodologies. The output of either 5C or 2C-ChIP protocols is a single-read sequencing library of ligated primer pairs that may or may not be multiplexed. While many computational tools currently exist for read mapping and analysis, these tools neither fully support multiplexed libraries nor provide qualitative reporting on the LMA primers involved. Typically, the task of library demultiplexing or primer analysis is offloaded on to the user. Our aim was to develop an easy-to-use pipeline for processing (multiplexed) single-read sequencing data produced by sequence-specific LMA.

**Results:** Here, we describe the Ligation-mediated Amplified, Multiplexed Primer-pair Sequence (LAMPS) analysis pipeline. LAMPS facilitates the analysis of multiplexed LMA sequencing data and provides a thorough assessment of a library's reads for a variety of experimental parameters (e.g., primer-pair efficiency). The standardized output of LAMPS allows for easy integration with downstream analyses, such as data track visualization on a genome browser. LAMPS is made publicly available on GitHub: <https://github.com/BlanchetteLab/LAMPS>

**Keywords:** Bioinformatics pipeline, Multiplexed ligation-mediated amplification, Carbon copy-chromatin immunoprecipitation, Chromosome Conformation Capture Carbon Copy

## Introduction

Ligation-Mediated Amplification (LMA) library preparation protocols [1, 2] have become increasingly useful in targeted sequencing methodologies. Applications include Carbon Copy-Chromatin Immunoprecipitation (2C-ChIP) [3], used to study protein-DeoxyriboNucleic Acid (DNA) interactions at a defined set of loci, and Chromosome Conformation Capture Carbon Copy (5C) [4], for the targeted analysis of chromatin architecture. These assays generally produce single-read sequencing data and are often highly multiplexed. The specificity

of the protocol makes it difficult to use generic bioinformatics pipelines for sequencing data analysis, such as Burrows-Wheeler Aligner's Smith-Waterman (BWA-SW) [5], Quantitative Insights Into Microbial Ecology (QIIME) 2 [6], or Torrent Mapping Alignment Program (TMAP) [7] (see Table 1). For example, the ligated, sequence-specific Forward (F) and Reverse (R) primers used in LMA may be incorrectly labeled as Polymerase Chain Reaction (PCR) duplicates or sequencing artifacts. Due to the low error rate of high-throughput sequencing and length of ligated primer-pair sequences, LMA reads are very similar and will map to the same genomic loci. In addition, analyses of LMA primer-pair (amplification) efficiencies for target DNA sequences are not standard. Users may also need to separately pre-process

\*Correspondence: christopher.cameron@mail.mcgill.ca

<sup>1</sup> School of Computer Science, McGill University, Montréal, Canada  
Full list of author information is available at the end of the article



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Table 1 Comparison of single-read mapping software**

	BWA-SW [5]	QIIME 2 [6]	TMAP [7]	LAMPS
Handles multiplexed reads	X	✓ <sup>a</sup>	X	✓
Primer QC reports provided	X	X	X	✓
2C-ChIP library normalization	X	X	X	✓
Genome browser formatted output	X	X	X	✓
Software dependencies	C Javascript	Python	C Python	Python Bowtie 2

<sup>a</sup> Using the 'q2-cutadapt' plugin

the sequencing data to demultiplex it. There exists a need for a computational pipeline that can more easily process LMA sequencing data, while providing diagnostic feedback on a variety of common issues that arise in this type of applications (i.e., primer-pair efficiency).

## Main text

### Materials and methods

Here we describe the Ligation-mediated Amplified, Multiplexed Primer-pair Sequence (LAMPS) analysis pipeline. LAMPS is a computational tool for mapping and analyzing sequence-specific LMA reads.

### Input

LAMPS takes as input a FAST-Quality (FASTQ) or Binary Alignment Map (BAM) file(s) obtained from the sequencing of a (possibly multiplexed) LMA-based library, together with a text file containing primer sequences and a configuration file describing optional normalization coefficients and barcode sequences.

### Mapping

LAMPS first uses either Bowtie 2 [8] (recommended) or Basic Local Alignment Search Tool (BLAST) [9] to map reads to the expected products containing all possible concatenations  $b \cdot p_1 \cdot p_2$ , where  $b$  is a BarCode (BC), and  $p_1$  and  $p_2$  each are either a F or R primer. Primer-pair counts are then tabulated for each BC (see matrix representation in Fig. 1). Reads that are either too short to map as a ligation product (i.e., less than the number of nucleotides to the primer-ligation junction of the shortest primer pair) or that do not map to the database (both cases termed 'Unmappable') are re-mapped to individual barcode BC-F (if needed), F, and R primer sequences for Quality Control (QC) reporting. QC reports are provided at both stages of mapping to identify underperforming primers and potential errors occurring within the protocol (i.e., human error, PCR artifacts, sequencing errors, etc.).

### Normalization

For each barcode, primer-pair read counts are normalized to Reads Per Million (RPM). When applicable (e.g., for 2C-ChIP libraries), tracks are normalized by input DNA counts and optionally corrected for sample-specific DNA density. Density correction is based on TaqMan quantification of total DNA yield following immunoprecipitation, as well as various dilution steps that occur in the preparation of pooled libraries as detailed in Wang and Cameron et al. [3].

### Output

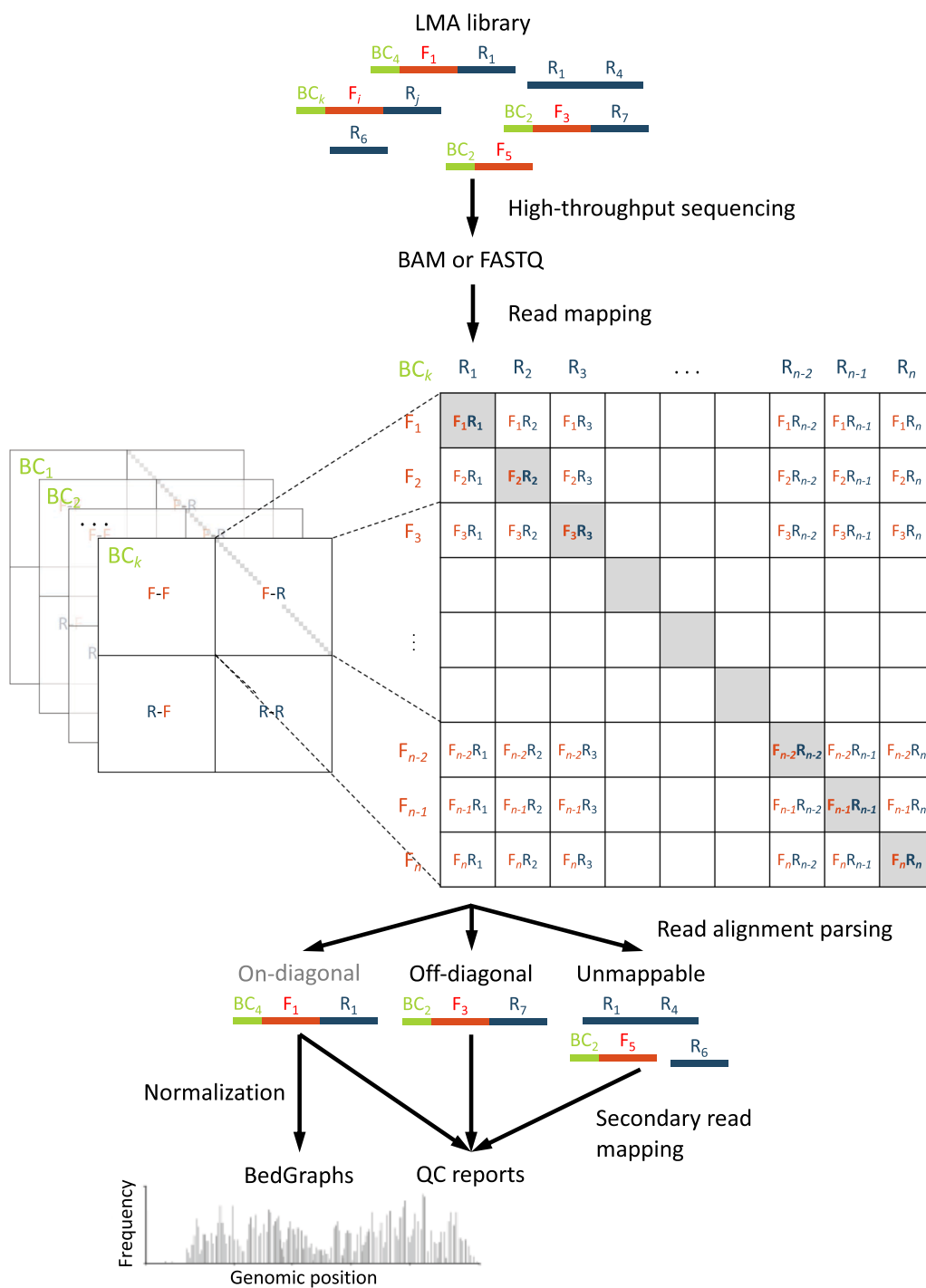
LAMPS's outputs depend on whether the experiment produces one- (1D) or two-Dimensional (2D) data (2C-ChIP and 5C, respectively). In the former case, raw and normalized primer-pair read counts are provided in bedGraph format for easy integration with most genomic browsers. For 2D data, raw interaction matrices at native resolution (e.g., that of individual restriction fragments for 5C) are provided.

Included within LAMPS's output are QC plots and reports to characterize the processed library. Library characterization includes, but is not restricted to, primer pair efficiency, raw and normalized read count comparison, and heatmaps describing the read count distribution.

### Implementation

LAMPS is a Unix-based (Linux and MacOS) command line pipeline, available in either Python v2 or v3 (versions 2.7.15 and 3.8.1 tested, respectively). Source code is available at: <https://github.com/BlanchetteLab/LAMPS>

The only dependencies of LAMPS are local installations of either the Bowtie 2 (recommended) or BLAST read aligner and (optional) Sequence Alignment Map tools (SAMtools) [10] (versions 2.3.4.2, 2.5.0+ and 1.3.1 tested, respectively). Included within the LAMPS GitHub repository are example 2C-ChIP and 5C datasets.



**Fig. 1** Schematic of the LAMPS analysis pipeline. LMA reads are mapped to the set of possible primer pairs (F-F, F-R, R-F, and R-R). If the sequenced library is multiplexed, the frequency count of each primer pair per barcode is obtained. Reads that are too short or could not be mapped initially ('Unmappable') are remapped to individual primer sequences. For 1D data, 'On-diagonal' read counts of expected primer pairs (gray entries of the F-R quadrant) are then normalized and outputted in bedGraph format. For 2D data, the entire F-R quadrant is provided as output in raw contact frequency matrix format

## Conclusion

LAMPS is a simple and easy-to-use computational tool for analyzing (multiplexed) sequence-specific LMA data. To the best of our knowledge, LAMPS is the first computational pipeline to provide thorough QC reporting of LMA primers. This reporting enables easy identification of problematic primer pairs during the design and data analysis of LMA experiments. To ensure LAMPS's ease of use, the pipeline natively handles multiplexed libraries that may result from a typical LMA protocol. In addition, the standardized format of LAMPS output allows data to be easily integrated with downstream analysis pipelines and quickly studied in the context of other genomic tracks.

## Limitations

LAMPS only corrects for known biases of 2C-ChIP data. Unknown primer biases that result from the 2C-ChIP protocol may contribute to erroneous results. LAMPS also does not perform fragment-bias normalization for 5C libraries and is expected to be run upstream of 5C bias-normalization pipelines. Finally, LAMPS is not currently available on Microsoft Windows. These limitations may be addressed in future iterations of the LAMPS software.

## Abbreviations

BAM: Binary Alignment Map; BC: BarCode; BLAST: Basic Local Alignment Search Tool; BWA-SW: Burrows-Wheeler Aligner's Smith-Waterman; DNA: DeoxyriboNucleic Acid; F: Forward primer; FASTQ: FAST-quality; LAMPS: Ligation-mediated Amplified, Multiplexed Primer-pair Sequence analysis pipeline; LMA: Ligation-Mediated Amplification; PCR: Polymerase Chain Reaction; QC: Quality control; QIIME: Quantitative Insights Into Microbial Ecology; R: Reverse primer; RPM: Reads Per Million; SAM: Sequence Alignment Map; TMAP: Torrent Mapping Alignment Program; 1D: One-Dimensional; 2C-ChIP: Carbon Copy-Chromatin Immunoprecipitation; 2D: Two-Dimensional; 5C: Chromosome Conformation Capture Carbon Copy.

## Acknowledgements

We would like to thank members of the Blanchette and Dostie laboratories for meaningful discussion during the development of this methodology and manuscript.

## Authors' contributions

CJFC developed the computational pipeline with the assistance of XQDW and supervision of JD and MB. CJFC and MB contributed to the paper writing. All authors have read and approved the final manuscript.

## Funding

This work was supported by the Canadian Institutes of Health Research (CIHR MOP-142451 to JD) and Natural Sciences and Engineering Research Council (NSERC Discovery grant to MB). XQDW was supported by a scholarship from CIHR and the Fonds de Recherche Santé Québec (FRQS).

## Availability of data and materials

LAMPS software and example input for 2C-ChIP and 5C libraries are made publicly available at <https://github.com/BlanchetteLab/LAMPS> and <https://doi.org/10.5281/zenodo.3858109> (GitHub repository and Zenodo archive, respectively).

## Ethics approval and consent to participate

Not applicable.

## Consent to publish

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup> School of Computer Science, McGill University, Montréal, Canada. <sup>2</sup> Department of Biochemistry and Goodman Cancer Research Center, McGill University, Montréal, Canada.

Received: 18 March 2020 Accepted: 21 May 2020

Published online: 03 June 2020

## References

- Guilfoyle RA, Leeck CL, Kroening KD, Smith LM, Guo Z. Ligation-mediated PCR amplification of specific fragments from a class-II restriction endonuclease total digest. *Nucleic Acids Res.* 1997;25(9):1854–8.
- Mueller PR, Wold B. Ligation-mediated PCR: applications to genomic footprinting. *Methods.* 1991;2(1):20–31.
- Wang XQD, Cameron CJF, Paquette D, Segal D, Warsaba R, Blanchette M, Dostie J. 2C-ChIP: measuring chromatin immunoprecipitation signal from defined genomic regions with deep sequencing. *BMC Genomics.* 2019;20:162.
- Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, Green RD, Dekker J. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 2006;16(10):1299–309.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60.
- Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope E, Da Silva R, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley G, Janssen S, Jarmusch AK, Jiang L, Kaehler B, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolk T, Kreps J, Langille MG, Lee J, Ley R, Liu Y, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton J, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson IMS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CH, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. QIIME 2: reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ Preprints.* 2018;6:27295–2.
- Homer N, Merriman B. TMAP: the Torrent Mapping Alignment Program. <https://github.com/iontorrent/TMAP>.
- Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357–9.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP. The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics.* 2009;25:2078–9.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.