

RESEARCH ARTICLE

Systematic Characteristic Exploration of the Chimeras Generated in Multiple Displacement Amplification through Next Generation Sequencing Data Reanalysis

Jing Tu[☉], Jing Guo[☉], Junji Li[☉], Shen Gao, Bei Yao, Zuhong Lu^{*}

State Key Lab of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing, China

☉ These authors contributed equally to this work.

* zhlu@seu.edu.cn



OPEN ACCESS

Citation: Tu J, Guo J, Li J, Gao S, Yao B, Lu Z (2015) Systematic Characteristic Exploration of the Chimeras Generated in Multiple Displacement Amplification through Next Generation Sequencing Data Reanalysis. PLoS ONE 10(10): e0139857. doi:10.1371/journal.pone.0139857

Editor: Junwen Wang, The University of Hong Kong, HONG KONG

Received: June 15, 2015

Accepted: September 16, 2015

Published: October 6, 2015

Copyright: © 2015 Tu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported project 61227803 and 61201343 of National Natural Science Foundation of China and Tsinghua National Laboratory for Information Science and Technology (TNList) Cross-discipline Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Background

The chimeric sequences produced by phi29 DNA polymerase, which are named as chimeras, influence the performance of the multiple displacement amplification (MDA) and also increase the difficulty of sequence data process. Despite several articles have reported the existence of chimeric sequence, there was only one research focusing on the structure and generation mechanism of chimeras, and it was merely based on hundreds of chimeras found in the sequence data of *E. coli* genome.

Method

We finished data mining towards a series of Next Generation Sequencing (NGS) reads which were used for whole genome haplotype assembling in a primary study. We established a bioinformatics pipeline based on subsection alignment strategy to discover all the chimeras inside and achieve their structural visualization. Then, we artificially defined two statistical indexes (the chimeric distance and the overlap length), and their regular abundance distribution helped illustrate of the structural characteristics of the chimeras. Finally we analyzed the relationship between the chimera type and the average insertion size, so that illustrate a method to decrease the proportion of wasted data in the procedure of DNA library construction.

Results/Conclusion

131.4 Gb pair-end (PE) sequence data was reanalyzed for the chimeras. Totally, 40,259,438 read pairs (6.19%) with chimerism were discovered among 650,430,811 read pairs. The chimeric sequences are consisted of two or more parts which locate inconsecutively but adjacently on the chromosome. The chimeric distance between the locations of adjacent parts on the chromosome followed an approximate bimodal distribution ranging

Competing Interests: The authors declare that they have no competing interests.

Abbreviations: MDA, multiple displacement amplification; WGA, whole genome amplification; NGS, next generation sequencing; SOAP2, Short Oligonucleotide Alignment Program 2; DOP-PCR, degenerate oligonucleotide primed polymerase chain reaction; PE, pair end; SE, single end; nt, nucleotide; bp, base pair.

from 0 to over 5,000 nt, whose peak was at about 250 to 300 nt. The overlap length of adjacent parts followed an approximate Poisson distribution and revealed a peak at 6 nt. Moreover, unmapped chimeras, which were classified as the wasted data, could be reduced by properly increasing the length of the insertion segment size through a linear correlation analysis.

Significance

This study exhibited the profile of the phi29MDA chimeras by tens of millions of chimeric sequences, and helped understand the amplification mechanism of the phi29 DNA polymerase. Our work also illustrated the importance of NGS data reanalysis, not only for the improvement of data utilization efficiency, but also for more potential genomic information.

Background

The whole genome amplification (WGA) methodologies have been developed for the purpose of human genome researches recently [1]. Initial WGA approaches were mainly based on the degenerate oligonucleotide primed polymerase chain reaction (DOP-PCR), which could amplify the product DNA several orders of magnitude than the initial template DNA [2, 3]. However, these approaches could lead to the nonspecific amplification and the amplification bias. Recent reports have described the use of non-PCR-based linear amplification protocols for WGA, and the bacteriophage phi29 DNA polymerase has been applied for the efficient amplification of circular DNA viral genomes without the need of specific primers in DOP-PCR [4, 5]. The phi29 DNA polymerase mediates an isothermal multiple displacement amplification (MDA) process, which omits the step to use the PCR instrument for the limiting initial DNA [6, 7]. It has high processivity and proofreading activity, and can generate large fragments (≥ 10 K bp) by using small template size (≥ 10 ng) with a good coverage rate and a reduced amplification bias [8]. Due to the obvious advantages, the phi29 DNA polymerase is being widely used in the procedure of microscale initial genomic DNA (gDNA) amplification in the WGA [9].

The chimeric sequence reads (*i.e.* chimeras) are consisted of two or more parts which locate inconsecutively but adjacently on the chromosome. In 2006, Zhang, K. *et al.* found chimeras in the whole genome shotgun sequence data of single-cell *Prochlorococcus* for genomic assembling, but the generation and structural details of chimerism lacked of enough elaboration [10]. Roger S Laken *et al.* used the 454 sequence platform to discover the phi29MDA chimeras in the sequence data of *E. coli* in 2007 [11]. However, this primary study was based on low-throughput sequence platform, and its analysis was not so explicit to explain the mechanism of phi29MDA chimera generation. Hundreds of chimeras (475 chimeras) was fairly far away from the significant level, which led to the lack of representativeness. Meanwhile, the genomic simplicity of the *E. coli* could not equivalently reflect the chimerism happening in human genome sequence data. With the increasingly usage of MDA in NGS, the phi29MDA chimerism could lead to severe disturbance in many studies, especially those focusing on heterogeneity and haplotype. Therefore, it is of great importance to elaborate the phi29MDA chimerism in WGA. From this standpoint, chimerism in human genome NGS data on the high-throughput sequence platform should be systematically studied.

Herein, we made further analysis focusing on phi29MDA human genome sequence data with the size over 130 Gb, which was used for whole genome haplotype assembling in 2013 [12]. For the purpose to systematically illustrate the regular patterns of the chimerism, we paid

close attention to all the non-pair-end alignment reads which were easier to be abandoned in conventional bioinformatics pipeline. Through our reanalysis to hundreds of millions of PE sequence reads, the results revealed the diversification of varied-level chimeras, and the statistically significant proportion of total chimeras in two samples (6.37% and 5.93% respectively, the detailed numbers were 24595400 and 15664038 PE read pairs respectively). Our study exhibited the profile of the phi29MDA chimerism through tens of millions of chimeras, and helped understand the amplification mechanism of the phi29 DNA polymerase.

Results

Data downloading and initial processing

The initial 206.8 Gb sequence data was downloaded from the Sequence Read Archive (SRA) database of National Center for Biotechnology Information (NCBI). According to the ethnic groups, the data was separated into two samples (SRX247249 and SRX252522). Since the existence of reads with N would influence the normalization of the following reanalysis, we filtered those reads with N, and 131.4 Gb raw reads was obtained after this procedure. The pair-end (PE) raw reads of each sample were aligned by Short Oligonucleotide Alignment Program 2 (SOAP2) in the PE alignment mode[13]. The data production was demonstrated as [Table 1](#).

Definition and visualization of the phi29MDA chimeras

The chimeras are consisted of two or more parts which locate inconsecutively but adjacently on the chromosome. Differing from the MDA primer polymer which are consisted of two or more sequence primers with the concordant length of 20–25 bp, the phi29MDA chimeras had at least one chimeric part whose length was definitely no shorter than 30 bp based on our bioinformatics pipeline ([Materials and Methods](#)), which meant the primer polymers would not mistakenly classified as the chimeras in our bioinformatics pipeline. On the aspect of the internal structure among the chimeric parts, there were overlap segments between the tail of the former part and the head of the following part. As for the chimeric strand specificity, this phenomena not only happened between two reverse strands, but also was observed on the same strand. Following the nomenclature of the previous research [11], fundamental chimerism could be divided into two types: inverted chimeras and direct chimeras. Two subsections of inverted chimeras are aligned to two reverse strands, while two subsections of direct chimera could be aligned concordantly to one strand. ([Fig 1](#)).

Proportion statistics of total and varied-level chimeras

Since the segment chimerism may happen for more than one times on a PE chimera read pair, it was necessary to classify the chimeras according to the levels of chimerism. Therefore, we generally separated the chimeras into three categories (1-level, 2-level and 3-level). For a 1-level chimera, only one time of chimerism happened, which lead to two sub-categories: the insertion chimera (the chimerism happened in the insertion segments) and the pair-end

Table 1. General data production.

Total Reads	Sample ID	Mapped	Unmapped
772587692(386293846 pairs)	SRX247249(101bp)	726523674(94.04%)	46064018(5.96%)
	SRX247249(30bp)	757652247(98.07%)	14935445(1.93%)
528273930(264136965 pairs)	SRX252522(101bp)	498760707(94.41%)	29513223(5.59%)
	SRX252522(30bp)	518496783(98.15%)	9777147(1.85%)

doi:10.1371/journal.pone.0139857.t001

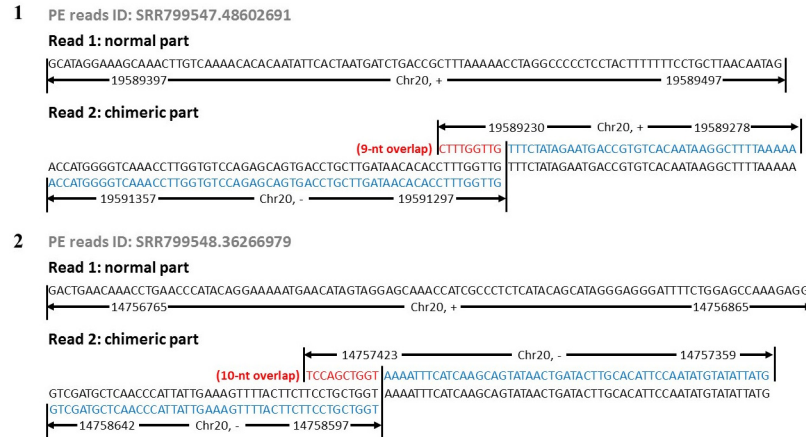


Fig 1. The visualization of the chimeras with the read parts located on the reverse strands (1) or on the same strands (2).

doi:10.1371/journal.pone.0139857.g001

chimera (the chimerism happened in the pair-end sequence reads). In similar ways, the 2-level and 3-level chimeras could be classified into several sub-categories. (Fig 2).

Totally, 24595400 (6.37%) and 15664038 (5.93%) PE read pairs were classified as chimeras in sample SRX247249 and SRX252522 respectively, which were both statistically significant. Moreover, the counting of the varied-level chimeras was finished according to the artificial classification of the chimeras above (Table 2). The results revealed that the majority of the chimeras (97.06% for SRX247249 and 95.71% for SRX252522 respectively) was in 1-level, and the higher-level chimeras were two orders of magnitude less than 1-level chimeras.

Since the segments of the chimeras could locate on either the same strands or the reverse strands, we also finished the quantity statistics of the pair-end chimeras, whose chimeric characteristics were visible, to see the related proportion of these two types (Fig 3). The results illustrated that the chimeric action could be easier to happen between two reverse strands (70.71% and 70.43% for each sample), which was generally in concordance with the results of previous researches [11].

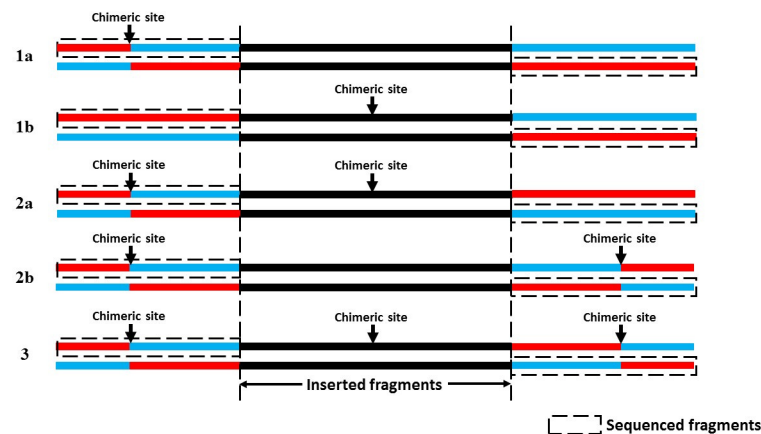


Fig 2. The characteristics of the sequence data of the varied-level chimeras. (1a) (1b) 1-level chimera; (2a) (2b) 2-level chimera; (3) 3-level chimera.

doi:10.1371/journal.pone.0139857.g002

Table 2. The amount of varied-level chimeras and the total proportion of chimeras.

Sample ID	Total read pairs	1-level chimeras	2-level chimeras	3-level chimeras	Total proportion
SRX247249	386293846	23872631	564397	158372	6.37%
SRX252522	264136965	14992729	445861	225448	5.93%

doi:10.1371/journal.pone.0139857.t002

Furthermore, according to our statistics, 74.55% (76.51% for SRX247249 and 71.47% for SRX252522 respectively, *i.e.* 4.61% of the total read pairs) of chimerism caused by phi29MDA in this series of sequence data happened in the insertion segments of pair-end reads. It led to the proportion increasement of the single-end alignment. However, the single-end mapped data is available for most ordinary bioinformatics researches. The remaining 25.45% chimeric read pairs (1.58% of the total read pairs) were the traditionally defined chimeras, which had at least one end with directly visual chimerism.

Distribution of the chimeric distance and the overlap length in 1-level chimeras

Theoretically, according to the enzyme kinetics of the phi29 DNA polymerase [14–16], as long as the overlap segments exists between two chimeric parts, meanwhile the chimeric distance was not very long, there was the probability to generate the chimeras. To assess the characteristics of the chimeras, we chose the chimeras with two chimeric parts on the reverse strands as the candidates, which had the largest number and the simplest structure. We analyzed the chimeric distance and the overlap length, and gave out the abundance distribution graph of the two samples (Fig 4). The results revealed that the abundance of the chimeric distance revealed an approximate bimodal distribution ranging from 0 to 5,000 nt, whose peak values was approximately 250 to 300 nt, *i.e.* the most probable chimeric distance. Furthermore, the abundance of the overlap length was observed as an approximate Poisson distribution with a peak of 6 nt, in which most chimeras had overlap segments ranging from 5 to 8 nt.

Relationship between the average insertion size and the chimera type

All the chimeras we obtained could be alternatively separated into two kinds: insertion chimeras and pair-end chimeras. Under normal circumstances, insertion chimeras could be regarded

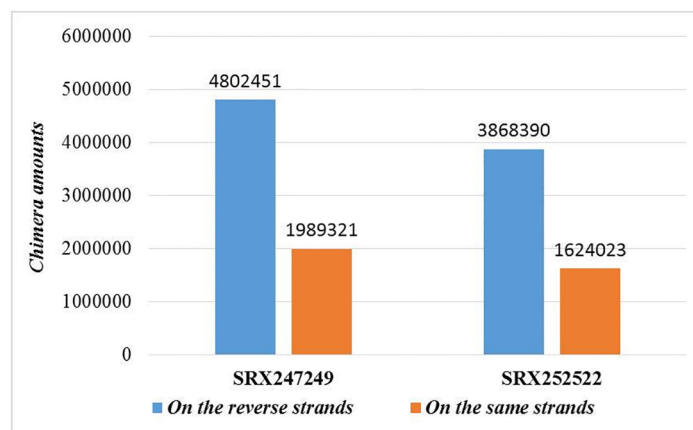


Fig 3. The quantity statistics of the pair-end chimeras with two parts on the reverse strands or on the same strands.

doi:10.1371/journal.pone.0139857.g003

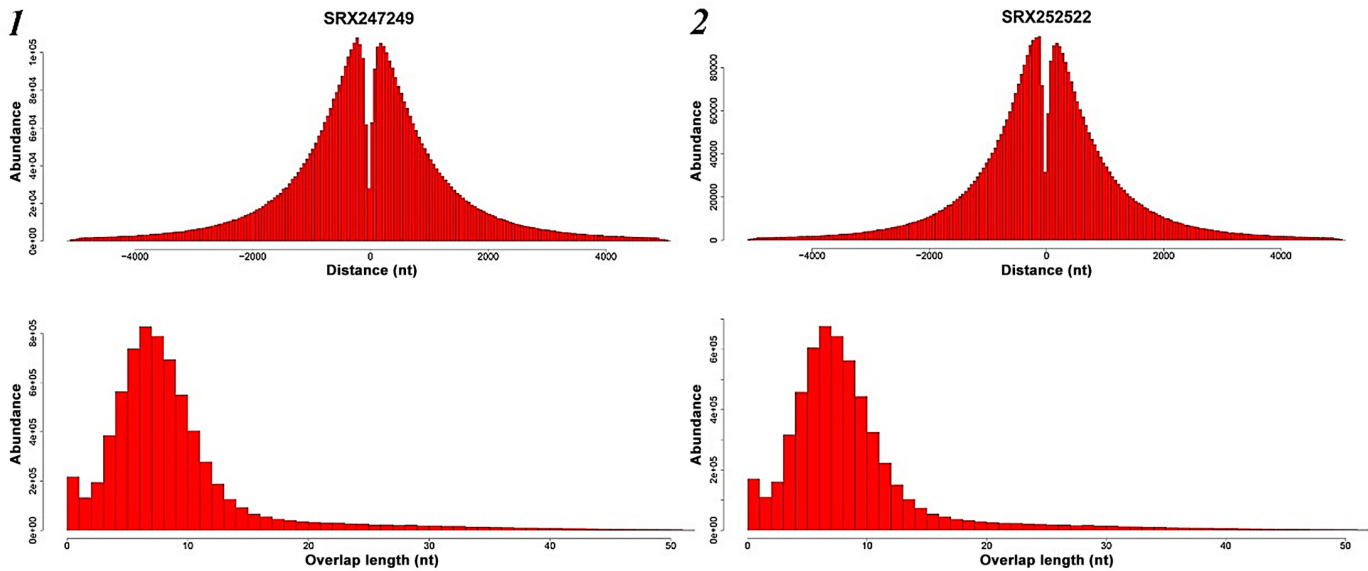


Fig 4. The abundance distribution graph of the physical genomic distance of the two parts of the reads and the length of the overlap segments. Sample SRX247249 was illustrated in Fig 4–1, and Sample SRX252522 was illustrated in Fig 4–2.

doi:10.1371/journal.pone.0139857.g004

as available data because their reads are completely mapped to the reference. However, the pair-end chimeras tend to be classified into the wasted data due to their inconformity with the genome. As the production of chimeras could not be avoided by using phi29MDA, it is necessary to investigate which factor directly influences the ratio of the insertion chimeras to the pair-end chimeras. For this purpose, we firstly hypothesized that the chimeric probability happening on each nucleotide of the PE reads was equally distributed. Since the sequenced length of the PE reads were constant (202 nt), theoretically the average insertion size was the principal parameter. The sequence data of the sample SRX247249 was divided into 18 parts according to the run number. Then their average insertion sizes and the ratios of the insertion chimeras to the pair-end chimeras were calculated (Table A in [S1 file](#)). The results revealed the significant positive correlation ($R^2 = 0.9691$) between the two indexes ([Fig 5](#)).

Discussion

In this study, we finished a further analysis towards the phi29MDA human genome sequence data of the whole genome haplotype assembling. We established a bioinformatics pipeline to comprehensively illustrate several aspects of the phi29MDA chimeras, *e.g.* the visualization of typical chimeras, the level classification of all varied chimeras, the proportion statistics of the varied-level chimeras, and the distribution statistics of two artificially defined indexes. The statistically significant total proportion of chimeras in WGA sequence data illustrated the importance to systematically study this special type of sequence data, which was easily ignored in genomic researches based on phi29MDA.

Normally, according to the different parameter settings in reads alignments, insertion chimeras could be regarded as wasted data in many researches using *de novo* sequencing, because they were classified as the single-end (SE) mapped reads in the standard SOAP2 alignment results [17]. Meanwhile, the pair-end chimeras were undoubtedly abandoned as wasted data in almost all researches using NGS. However, these chimeras could be translated into multiple-fragments available data through our analytical pipeline. Therefore, our study contributed to the improving of the sequence data utilization efficiency.

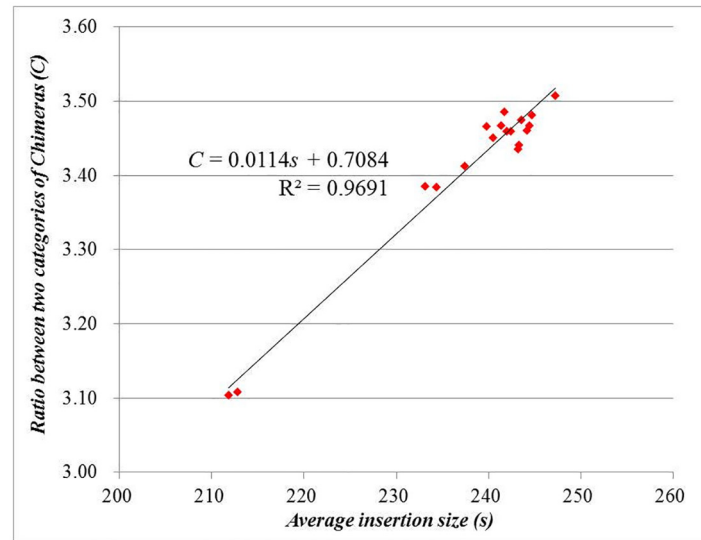


Fig 5. The curve of the 18 subsamples about the relationship between the average insertion size and the ratio of the insertion chimeras to the pair-end chimeras.

doi:10.1371/journal.pone.0139857.g005

During the process of phi29 MDA, strand displacements are happening with a high frequency, which leads to the construction of the hyperbranched structure. Herein, due to the ultrahigh complexity and scale of the human genome, almost all DNA polymers are always “facing” countless binding positions and would bind to one of them which could keep the polymers in most energetically stable condition. According to this inference, DNA polymers definitely tend to choose one close position with relatively long overlap segments generally. Therefore, inverted chimeras are easier to be generated according to its relatively simpler structure; on the contrary, direct chimeras are more difficult to be generated thermodynamically because of the complex structure. On the same principle, low-level chimeras are easier to be generated (Figure C in [S1 file](#)).

Since the insertion chimeras are directly available while the pair-end chimeras could be easily classified as unmapped reads, it is necessary to discover the influencing factor towards the ratio between the two categories of chimeras. We used data of the sample SRX247249 to construct 18 subsamples according to the run numbers, in order to see the relationship between the average insertion sizes and the ratios between the two categories of chimeras, and found the significant positive correlation between two indexes. The results revealed that proper increasement of the average insertion size in standard DNA library construction for Illumina Sequence Platform was effective for the proportion reduction of the pair-end chimeras. As for the aspect of bioinformatics analysis, after being cut into two or more subsections according to the chimeric sites through our pipeline, the segments of chimeras could be integrated into a new FASTQ file, which would be aligned to the reference genome again. Thus, the chimeras are recycled instead of being thrown into the unmapped wasted data. By comprehensively using both experimental and bioinformatics solutions, researchers are able to increase the sequence data utilization efficiency with this strategy.

In comparison with previous researches about phi29MDA chimeras, reanalyzed data in our study was sequenced from human genome, which is more complex structurally and dimensionally than genomes of prokaryotes. Due to the extreme complicacy of the human genome, the generation of a great quantity of chimeras in DNA isothermal amplification may not only reflect the genomic revolution from simple prokaryotes to complicated Homo Sapiens, but also

be persuasive to explain the general sequence symmetry of human genome, *i.e.* the Chargaff 2nd Parity Laws [18].

In spite of the chimera generation, phi29 DNA polymerase is undoubtedly worth being used in many WGA experiments based on its obvious advantages. Our study broadens the horizons about the enzyme kinetics of the phi29 DNA polymerase, and may be helpful for the comprehensive explanation of the thermodynamics of the phi29MDA process. Furthermore, the unique multiple-fragments structure of the chimeras illustrates the potential value in future studies of human genome [19].

Materials & Methods

Data Sources

Genome sequence data was downloaded from National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database, and the SRA numbers were SRX247249 and SRX252522. The type of the sequence data was PE-101.

Bioinformatics pipeline establishment

The phi29MDA chimeras could be divided into two categories: the insertion chimeras and the pair-end chimeras. Since we could find the insertion chimeras by directly observing the mapping locations (+ or—strands) of the PE reads and judging their logical relation, we established a bioinformatics pipeline based on the subsection alignment strategy for the purpose of the detailed analysis to the pair-end chimeras. The pipeline could be mainly divided into two parts: A. reads alignment and filtering; B. detailed analysis of the candidate chimeras.

A. Reads alignment and filtering. Firstly we obtained the PE-101 clean reads after removal of the PE-101 raw reads with N. The clean reads would be conducted in two ways. On one hand, they were mapped to the Hg19 human genome reference by using SOAP2 software in PE mode. From the alignment output data, we abandoned the PE mapped reads, and kept the other reads (including SE mapped reads and unmapped reads) for the following analysis. On the other hand, we cut 30-nt seeds from the clean data, and constructed a series of PE-30 clean reads. These reads were also mapped to the Hg19 human genome reference by using SOAP2 software in PE mode. From the alignment output data, we abandoned the unmapped reads, and kept the other reads (including PE and SE mapped reads) for the following analysis. According to the reads ID, we found the PE-101 reads from the clean data with the characteristics described as: their 30-nt seeds could accurately mapped to the genome, while they could not on the same location. These reads would be collected as the candidate reads for the following detailed analysis about the pair-end chimeras (Figure A in [S1 file](#)).

B. Subsection alignment strategy used for chimera search. In this step, we aimed to accomplish the discovery of all the pair-end chimeras, meanwhile understand their exact structure. This step mainly included two core substeps: seed extension and local alignment. Initially, the mapped 30-nt seeds were extended one nucleotide by one nucleotide on their related whole reads until they arrived at a mismatch site. Here we obtained the two subsections of the pair-end chimeras and the start-end coordinates of the former subsections on the reference genome. Afterwards, the following subsections of the pair-end chimeras were locally aligned in the 5,000 bp local region both upstream and downstream from the end coordinate of the former subsections in order to find their exact location in the genome. Thus we could know start-end coordinates of the following subsections of the pair-end chimeras. Finally, we reversely extended the following subsections one nucleotide by one nucleotide from their start coordinates on the reference genome, and compared the nucleotide type with the end of the former subsections, in order to obtain the overlap segments (Figure B and Figure D in [S1 file](#)).

According to the start-end coordinates of the two subsections of the pair-end chimeras, we could calculate the chimeric distance through the equation:

$$D = C_{\text{former, end}} - C_{\text{following, start}}$$

In the equation, D meant the chimeric distances, $C_{\text{former, end}}$ meant the end coordinates of the former subsections, and $C_{\text{following, start}}$ meant the start coordinates of the following subsections.

Supporting Information

S1 File. Supporting figures and tables.
(PDF)

Acknowledgments

This work was supported project 61227803 and 61201343 of National Natural Science Foundation of China and Tsinghua National Laboratory for Information Science and Technology (TNList) Cross-discipline Foundation. We thank Shengqin Wang for the guidance about the data processing.

Author Contributions

Conceived and designed the experiments: JT ZL. Analyzed the data: JG JT JL SG BY. Wrote the paper: JT JG. Designed the analysis pipeline, drafted and revised the manuscript: JG JT JL. Participated in programming: SG. Participated in result analyzing: BY. Read and approved the final manuscript: JT JG JL SG BY ZL.

Reference

1. Hughes S, Arneson N, Done S, Squire J. The use of whole genome amplification in the study of human disease. *Progress in biophysics and molecular biology*. 2005; 88(1):173–89. doi: [10.1016/j.pbiomolbio.2004.01.007](https://doi.org/10.1016/j.pbiomolbio.2004.01.007) PMID: [15561304](https://pubmed.ncbi.nlm.nih.gov/15561304/).
2. Telenius Hk, Carter NP, Bebb CE, Nordenskjöld M, Ponder BAJ, Tunnacliffe A. Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer. *Genomics*. 1992; 13(3):718–25. doi: [10.1016/0888-7543\(92\)90147-K](https://doi.org/10.1016/0888-7543(92)90147-K) PMID: [1639399](https://pubmed.ncbi.nlm.nih.gov/1639399/)
3. Kittler R, Stoneking M, Kayser M. A Whole Genome Amplification Method to Generate Long Fragments from Low Quantities of Genomic DNA. *Analytical Biochemistry*. 2002; 300(2):237–44. doi: [10.1006/abio.2001.5460](https://doi.org/10.1006/abio.2001.5460) PMID: [11779116](https://pubmed.ncbi.nlm.nih.gov/11779116/)
4. John R, Muller H, Rector A, van Ranst M, Stevens H. Rolling-circle amplification of viral DNA genomes using phi29 polymerase. *Trends in microbiology*. 2009; 17(5):205–11. doi: [10.1016/j.tim.2009.02.004](https://doi.org/10.1016/j.tim.2009.02.004) PMID: [19375325](https://pubmed.ncbi.nlm.nih.gov/19375325/).
5. Dean FB, Nelson JR, Giesler TL, Lasken RS. Rapid Amplification of Plasmid and Phage DNA Using Phi29 DNA Polymerase and Multiply-Primed Rolling Circle Amplification. *Genome Res*. 2001; 11(6):1095–9. doi: [10.1101/gr.180501](https://doi.org/10.1101/gr.180501) PMC311129. PMID: [11381035](https://pubmed.ncbi.nlm.nih.gov/11381035/)
6. Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat Genet*. 1998; 19(3):225–32. PMID: [9662393](https://pubmed.ncbi.nlm.nih.gov/9662393/)
7. Shoaib M, Baconnais S, Mechold U, Le Cam E, Lipinski M, Ogryzko V. Multiple displacement amplification for complex mixtures of DNA fragments. *BMC Genomics*. 2008; 9:415-. doi: [10.1186/1471-2164-9-415](https://doi.org/10.1186/1471-2164-9-415) PMC2553422. PMID: [18793430](https://pubmed.ncbi.nlm.nih.gov/18793430/)
8. Paez JG, Lin M, Beroukhim R, Lee JC, Zhao X, Richter DJ, et al. Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Res*. 2004; 32(9):e71–e. doi: [10.1093/nar/gnh069](https://doi.org/10.1093/nar/gnh069) PMC419624. PMID: [15150323](https://pubmed.ncbi.nlm.nih.gov/15150323/)
9. Han T, Chang C-W, Kwekel JC, Chen Y, Ge Y, Martinez-Murillo F, et al. Characterization of whole genome amplified (WGA) DNA for use in genotyping assay development. *BMC Genomics*. 2012; 13:217-. doi: [10.1186/1471-2164-13-217](https://doi.org/10.1186/1471-2164-13-217) PMC3403925. PMID: [22655855](https://pubmed.ncbi.nlm.nih.gov/22655855/)

10. Zhang K, Martiny AC, Reppas NB, Barry KW, Malek J, Chisholm SW, et al. Sequencing genomes from single cells by polymerase cloning. *Nature biotechnology*. 2006; 24(6):680–6. doi: [10.1038/nbt1214](https://doi.org/10.1038/nbt1214) PMID: [16732271](https://pubmed.ncbi.nlm.nih.gov/16732271/).
11. Lasken RS, Stockwell TB. Mechanism of chimera formation during the Multiple Displacement Amplification reaction. *BMC biotechnology*. 2007; 7:19. doi: [10.1186/1472-6750-7-19](https://doi.org/10.1186/1472-6750-7-19) PMID: [17430586](https://pubmed.ncbi.nlm.nih.gov/17430586/); PubMed Central PMCID: PMC1855051.
12. Kaper F, Swamy S, Klotzle B, Munchel S, Cottrell J, Bibikova M, et al. Whole-genome haplotyping by dilution, amplification, and sequencing. *Proc Natl Acad Sci U S A*. 2013; 110(14):5552–7. doi: [10.1073/pnas.1218696110](https://doi.org/10.1073/pnas.1218696110) PMC3619281. PMID: [23509297](https://pubmed.ncbi.nlm.nih.gov/23509297/)
13. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, et al. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*. 2009; 25(15):1966–7. doi: [10.1093/bioinformatics/btp336](https://doi.org/10.1093/bioinformatics/btp336) PMID: [19497933](https://pubmed.ncbi.nlm.nih.gov/19497933/).
14. Perez-Arnaiz P, Longas E, Villar L, Lazaro JM, Salas M, de Vega M. Involvement of phage phi29 DNA polymerase and terminal protein subdomains in conferring specificity during initiation of protein-primed DNA replication. *Nucleic Acids Res*. 2007; 35(21):7061–73. doi: [10.1093/nar/gkm749](https://doi.org/10.1093/nar/gkm749) PMID: [17913744](https://pubmed.ncbi.nlm.nih.gov/17913744/); PubMed Central PMCID: PMC2175359.
15. Rodriguez I, Lazaro JM, Salas M, de Vega M. Involvement of the TPR2 subdomain movement in the activities of phi29 DNA polymerase. *Nucleic Acids Res*. 2009; 37(1):193–203. doi: [10.1093/nar/gkn928](https://doi.org/10.1093/nar/gkn928) PMID: [19033368](https://pubmed.ncbi.nlm.nih.gov/19033368/); PubMed Central PMCID: PMC2615600.
16. Morin JA, Cao FJ, Lazaro JM, Arias-Gonzalez JR, Valpuesta JM, Carrascosa JL, et al. Active DNA unwinding dynamics during processive DNA replication. *Proc Natl Acad Sci U S A*. 2012; 109(21):8115–20. doi: [10.1073/pnas.1204759109](https://doi.org/10.1073/pnas.1204759109) PMID: [22573817](https://pubmed.ncbi.nlm.nih.gov/22573817/); PubMed Central PMCID: PMC3361432.
17. Chitsaz H, Yee-Greenbaum JL, Tesler G, Lombardo MJ, Dupont CL, Badger JH, et al. Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. *Nature biotechnology*. 2011; 29(10):915–21. doi: [10.1038/nbt.1966](https://doi.org/10.1038/nbt.1966) PMID: [21926975](https://pubmed.ncbi.nlm.nih.gov/21926975/); PubMed Central PMCID: PMC3558281.
18. Wang S, Tu J, Jia Z, Lu Z. High order intra-strand partial symmetry increases with organismal complexity in animal evolution. *Scientific reports*. 2014; 4:6400. doi: [10.1038/srep06400](https://doi.org/10.1038/srep06400) PMID: [25263801](https://pubmed.ncbi.nlm.nih.gov/25263801/); PubMed Central PMCID: PMC4178289.
19. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *Journal of computational biology: a journal of computational molecular cell biology*. 2013; 20(10):714–37. doi: [10.1089/cmb.2013.0084](https://doi.org/10.1089/cmb.2013.0084) PMID: [24093227](https://pubmed.ncbi.nlm.nih.gov/24093227/); PubMed Central PMCID: PMC3791033.