ERRATUM

From Sequence Mapping to Genome Assemblies

Thomas D. Otto

Christopher Peacock (ed.), *Parasite Genomics Protocols*, Methods in Molecular Biology, vol. 1201,
DOI 10.1007/978-1-4939-1438-8_2, © Springer Science+Business Media New York 2015

DOI 10.1007/978-1-4939_1438-8_21

Figures 1, 2 and 3 of this chapter is incorrect. The correct figures are as shown below.

Also there are 7 additional references that were missed to be added in the final version of the book. They are as listed below.

CITATION 1
Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species.

Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S, Chapman JA, Chapuis G, Chikhi R, Chitsaz H, Chou WC, Corbeil J, Del Fabbro C, Docking TR, Durbin R, Earl D, Emrich S, Fedotov P, Fonseca NA, Ganapathy G, Gibbs RA, Gnerre S, Godzaridis E, Goldstein S, Haimel M, Hall G, Haussler D, Hiatt JB, Ho IY, Howard J, Hunt M, Jackman SD, Jaffe DB, Jarvis ED, Jiang H, Kazakov S, Kersey PJ, Kitzman JO, Knight JR, Koren S, Lam TW, Lavenier D, Laviolette F, Li Y, Li Z, Liu B, Liu Y, Luo R, Maccallum I, Macmanes MD, Maillet N, Melnikov S, Naquin D, Ning Z, Otto TD, Paten B, Paulo OS, Phillippy AM, Pina-Martins F, Place M, Przybylski D, Qin X, Qu C, Ribeiro FJ, Richards S, Rokhsar DS, Ruby JG, Scalabrin S, Schatz MC, Schwartz DC, Sergushichev A, Sharpe T, Shaw TI, Shendure J, Shi Y, Simpson JT, Song H, Tsarev F, Vezzi F, Vicedomini R, Vieira BM, Wang J, Worley KC, Yin S, Yiu SM, Yuan J, Zhang G, Zhang H, Zhou S, KorfIF.Gigascience. 2013 Jul 22;2(1):10. doi: 10.1186/2047-217X-2-10.
PMID:23870653[PubMed]

The online version of the original chapter can be found at
http://dx.doi.org/10.1007/978-1-4939-1438-8_2

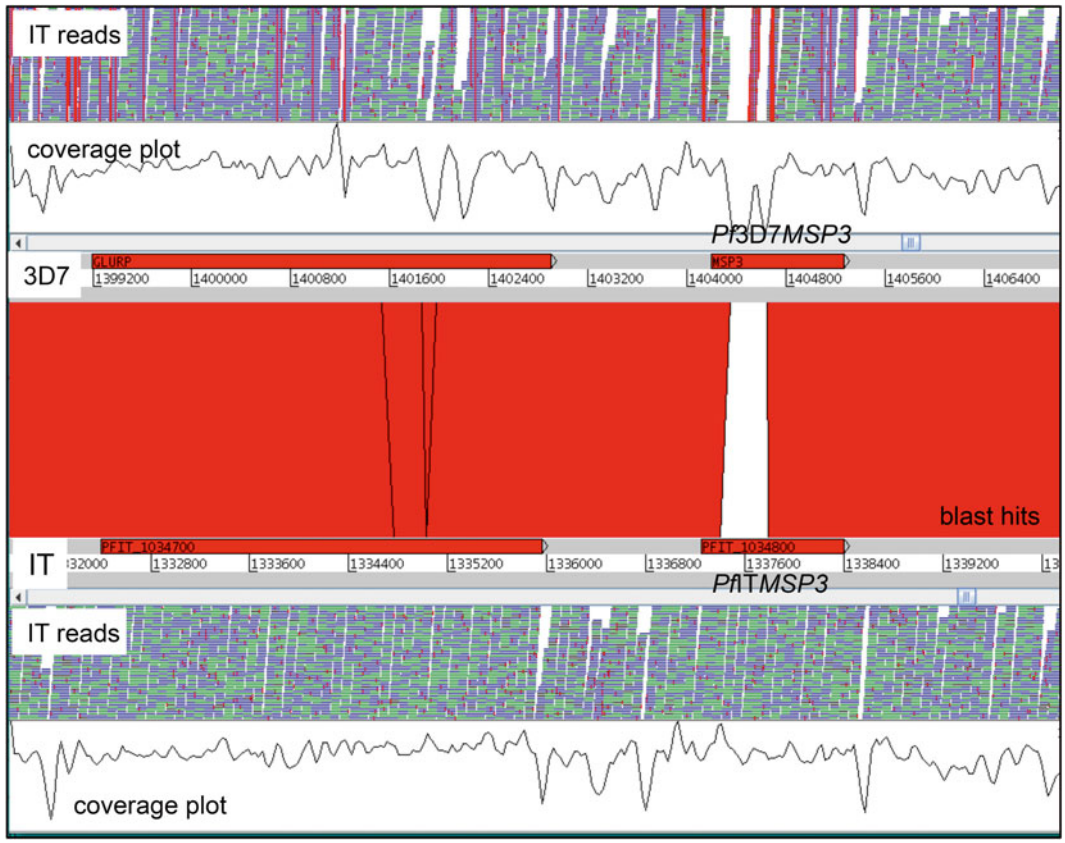

Fig. 1 Mapping versus assembly. Two genes of *P. falciparum* 3D7 (*red boxes*) can be seen at the top. The *horizontal green* and *blue lines* are mapped sequencing reads from the IT clone. *Red points* in the reads are differences between the IT reads and the 3D7 reference. The lower part shows the *de novo* assembly of IT. The *vertical bars* are blast hits. The graphs are the coverage plots. Some regions of *MSP3* in 3D7 are not covered by mapped IT reads. The de novo assembly has an insertion, indicated by the shape of the blast hit. Reads map even over this new assembled region

CITATION 2

A comprehensive evaluation of assembly scaffolding tools.Hunt M, Newbold C, Berriman M, Otto TD.Genome Biol. 2014 Mar 3;15(3):R42. [Epub ahead of print]PMID:24581555

CITATION 3

REAPR: a universal tool for genome assembly evaluation.Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto TD.Genome Biol. 2013 May 27;14(5):R47. doi: 10.1186/gb-2013-14-5-r47.PMID:23710727

CITATION 4

A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and IlluminaMiSeq sequencers.
Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y.
BMC Genomics. 2012 Jul 24;13:341. doi: 10.1186/1471-2164-13-341.PMID:22827831

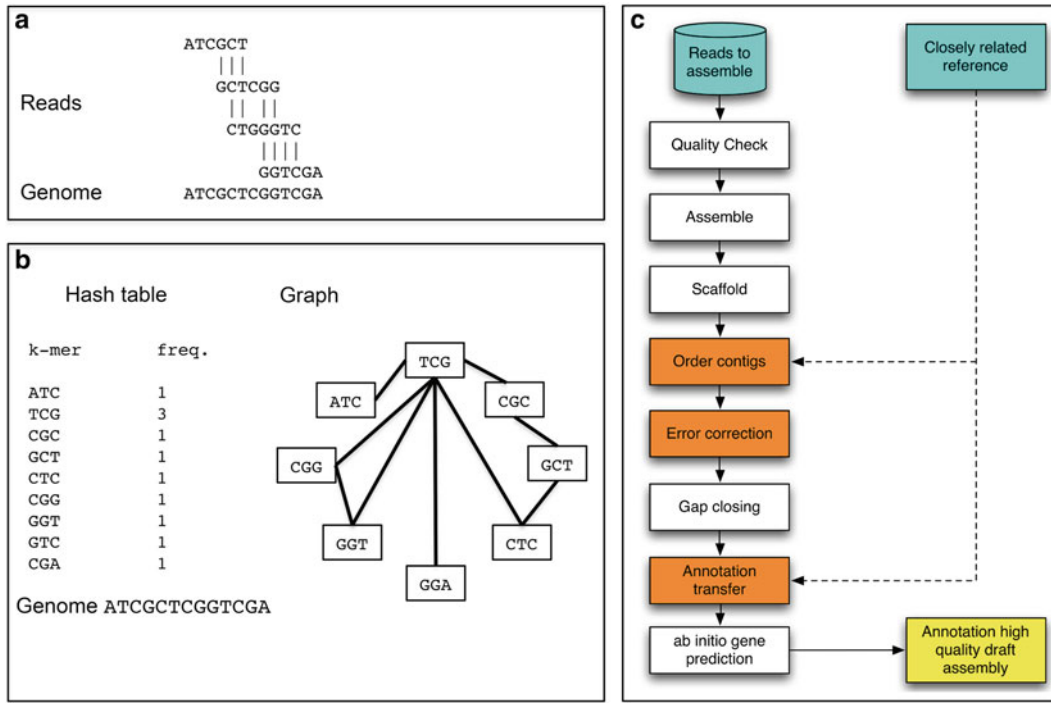

Fig. 2 (**a**) Assembly with longer reads: Nearly identical overlap between reads enable the generation of the consensus. (**b**) Assembly with short reads, using de Bruijn graph: First the reads are index and the k-mer are stored in a hash table, including the k-mer and the frequency. With a k-mer length of 3 the k-mer TCG is non unique. Due to this non unique k-mer, the graph quite complicated. (**c**) Overview of typical pipeline for de novo assembly and annotation

CITATION 5 and 6

- Iddo Friedberg Automated protein function prediction—the genomic challenge Brief Bioinform (2006) 7 (3): 225-242 first published online May 23, 2006 doi:10.1093/bib/bbl004

Nat Rev Genet. 2012 Apr 18;13(5):329-42. doi: 10.1038/nrg3174.

A beginner's guide to eukaryotic genome annotation.

Yandell M1, Ence D.

CITATON 7 (prokka)

Bioinformatics. 2014 Jul 15;30(14):2068-9. doi: 10.1093/bioinformatics/btu153. Epub 2014 Mar 18.

Prokka: rapid prokaryotic genome annotation.

Seemann T.PMID:

24642063

[PubMed - in process]

a

```
>Pf3D7_10_v3
taaaccctgaaccctaaaccctgaaccctaaaccactaaccctaaaccctgaaccctgaa
ccctgaaccctaaaccctaaacctaaaccctgaaccctaaaacctaaaccctgaaccctg
aaccctaaaccctaaaccctaaaccctaaaccctgaaccctaaaccctaaaccctgaacc
ctaaaccctgaaccctgaaccctaaaccctgaaccctgaaccctaaaccctgaaccacta
```

b

```
@IL39_6014:8:110:3699:4595#3/1
TATTTGAACTGACAATTTTATAAGATCCATATATATGAAGATCTCAAAAAAATATATGTTTTTTTGAAAATTTTCA
+
CCB@@*CCC<GCHHEHGGGGGEEGDDGGGGGEGGHEGCEGGFFCHGDBCB?BH?GGBH=BEBB@@E=B>=EBECGA
@IL39_6014:8:88:4857:8768#5/1
TATTTAAACTGACAATTTTATAAGATCCATATATATGAAGATCTCAAAAAAATATATGTTTTTTTGAAAATTTTCA
```

c

```
IL39_6014:8:110:3699:4595#3   83        PfIT_10_v2          1404405   60        76M
          =          1404336   -145
          TGAAAATTTTCAAAAAAACATATATTTTTTTGAGATCTTCATATATATGGATCTTATAAAATTGTCAGTTCAAATA
          AGCEBE=>B=E@@BBEB=HBGG?HB?BCBDGHCFFGGECGEHGGEGGGGGDDGEEGGGGGHEHHCG<CCC*@@BCC
          AS:i:73
```

Fig. 3 Examples of different file formats. (**a**) fasta: Each sequence starts with a ">" and a name. Then the sequence is followed. (**b**) fastq: Similar to fasta, but with the quality coded in ASCII. (**c**) SAM format: First column is the name of the read. Next column is the mapping flag that can be used for querying a BAM file. Third and fourth, seven and eight columns are mapped to the reads and its mate, respectively. Column nine is the fragment size. The information how well the reads map is in column five and six, mapping quality and cigar string, respectively. The sequence and the quality of the reads are stored in column ten and eleven. The last column can have many different information, like an alignment score, other possible position to map repetitively. This depends on the mapper

CITATION 8 trimmomatic

Bioinformatics. 2014 Aug 1;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. Epub 2014 Apr 1.

Trimmomatic: a flexible trimmer for Illumina sequence data.

Bolger AM1, Lohse M2, Usadel B1.

PMID:

24695404