Updating benchtop sequencing performance comparison

To the Editor:

In April 2012, your journal published a study by Loman *et al.*[1] that systematically compared desktop next-generation sequencers (NGS) from three instrument providers. Using the custom scripts supplied by the authors, the same software and the same draft genome (with 153 remaining gaps within several scaffolds) as the reference, we reproduced their results with their data of the enterohemorrhagic *Escherichia coli* (EHEC) strain found in the 2011 outbreak in Germany. However, we wish to bring readers' attention to some shortcomings in the report from Loman *et al.*[1], focusing particularly on its discussion of read-level error analysis.

NGS is a rapidly changing market, which clearly complicates the comparisons such as that made by Loman *et al.* Since the original study[1], Illumina (San Diego) has launched the MiSeq sequencer officially and has released Nextera library construction kits and 2 × 250–base-pair (250-bp) paired-end (PE) sequencing chemistry. Furthermore, Life Technologies (Carlsbad, California), has made 200-bp and 300-bp kits available for the Ion Torrent Personal Genome Machine (PGM). Roche (Basel, Switzerland) has updated the Sequencing System software for its 454 GS Junior (GSJ) from version 2.6 to 2.7. In this report, we provide an up-to-date snapshot of how benchtop platforms have evolved since the previous study[1].

To assess accuracy and the contiguity of draft assemblies on a finished genome, we based our analysis on the finished sequence of the enterohemorrhagic *E. coli* O157:H7 Sakai strain[2]. All Sakai DNA used in this study was prepared from the same subcultivation. Aliquots of this DNA were shipped to three academic institutions for whole-genome sequencing on the GSJ, MiSeq and PGM. The three institutions chosen are successful operators of the respective instruments. Thus, no NGS platform manufacturer was involved with this study. For all three platforms, the latest available software and most recent chemistries (that is, the GSJ Titanium, the MiSeq Nextera library with 2 × 250-bp PE and the PGM 300-bp kit) were applied.

By assembling sequencing reads against the existing Sakai backbone genome (**Supplementary Methods**), we generated run and mapping metrics for each benchtop sequencer (**Supplementary Table 1**). Of the three instruments, the GSJ produced the lowest throughput, which was insufficient for assembling typical bacterial genomes. Therefore, we combined two GSJ runs into a single data set for all subsequent analyses. The relationship between chromosomal and plasmid (large plasmid pO157 and small plasmid pOSAK1) average coverage was similar for all data. The coverage along the genome was even for all technologies (data not shown). GSJ produced the longest reads, with a mean length of 466 bases. The mean read length increased for MiSeq from 142 bases (2 × 150-bp PE) to 214 bases (2 × 250-bp PE) and for PGM from 116 bases (100-bp kit) to 195 bases (300-bp kit). Comparison of error frequencies on read level per 100 bp showed that rates of insertion and deletion (indel) and substitution for MiSeq 2 × 150-bp PE and GSJ were similar to those reported by Loman *et al.*[1] (**Table 1**). However, we also observed a fourfold decreased substitution rate and a threefold lower indel rate for the 100-bp PGM, compared with previous results from Loman *et al.*[1]. The error profile was unchanged by the use of longer (2 × 250-bp PE) MiSeq read lengths. In contrast, the rate of indels in the PGM data doubled after the introduction of a new enzymatic formulation with the 300-bp chemistry. For the newest chemistries, all three platforms had substitution error rates of the same order of magnitude, whereas the MiSeq clearly had the lowest number of indels per read.

To detect differences in gene content using NGS, one must generate accurate *de novo* assemblies. *De novo* assemblers combine reads to create full-length contiguous sequences without the guidance of a reference genome. We obtained *de novo* assembly metrics by applying the overlap-based genome assembler MIRA (version 3.4.0). Assembly efficiency is greatly influenced by the number of reads covering the entire genome. Furthermore, with higher coverage, the computational effort increases substantially. Therefore, we first evaluated how increasing coverage affects N50 contig size (a statistic for describing the distribution of contig lengths in an assembly)[3] for the three platforms with their newest available chemistries (**Supplementary Fig. 1**). For MiSeq, a coverage of >75-fold did not yield further improvement in terms of N50. This is consistent with results obtained using simulated 75-bp PE data and the Velvet assembler[4]. The optimal coverage using data from the PGM instrument was ~40-fold. In contrast to results obtained using data from the MiSeq, the N50 size of assemblies based on data from the PGM decreased after reaching the optimal coverage. For the GSJ, even combined data from two runs did not yield optimum or plateau N50 values. Therefore, all further *de novo* assembly metrics were computed with data randomly subsampled to 75-fold coverage for the MiSeq, 40-fold coverage for the PGM and the two combined GSJ runs. When comparing GSJ data with MiSeq 2 × 150-bp PE and PGM 100-bp or 200-bp data, we obtained assemblies from the GSJ data that were less fragmented, owing to the longer read

Table 1 Insertion/deletion and substitution errors on read level for benchtop NGS platforms

Platform	Sequencing kit	Library	Strain	Date of sequencing	Indels per 100 bp	Indels per read	Substitutions per 100 bp	Substitutions per read
GSJ	GSJ Titanium	Nebulization / AMPure XP	Sakai	June 2012	0.4011	1.8351	0.0543	0.2484
MiSeq	2 × 150-bp PE	Nextera	Sakai	June 2012	0.0009	0.0013	0.0921	0.1318
MiSeq	2 × 250-bp PE	Nextera	Sakai	September 2012	0.0009	0.0018	0.0940	0.2033
PGM	100 bp	Bioruptor / Ion Fragment Library	Sakai	July 2011	0.3520	0.3878	0.0929	0.1024
PGM	200 bp	Ion Xpress Plus Fragment	Sakai	July 2012	0.3955	0.6811	0.0303	0.0521
PGM	300 bp	Ion Xpress Plus Fragment	Sakai	August 2012	0.7054	1.4457	0.0861	0.1765
PGM	400 bp[a]	Ion Xpress Plus Fragment	Sakai	November 2012	0.6722	1.8726	0.0790	0.2202

Error rates were calculated by counting indels and substitutions in the mapping against the EHEC Sakai reference sequence for each uniquely mapped read.
[a]Kit was not officially available during time of study.

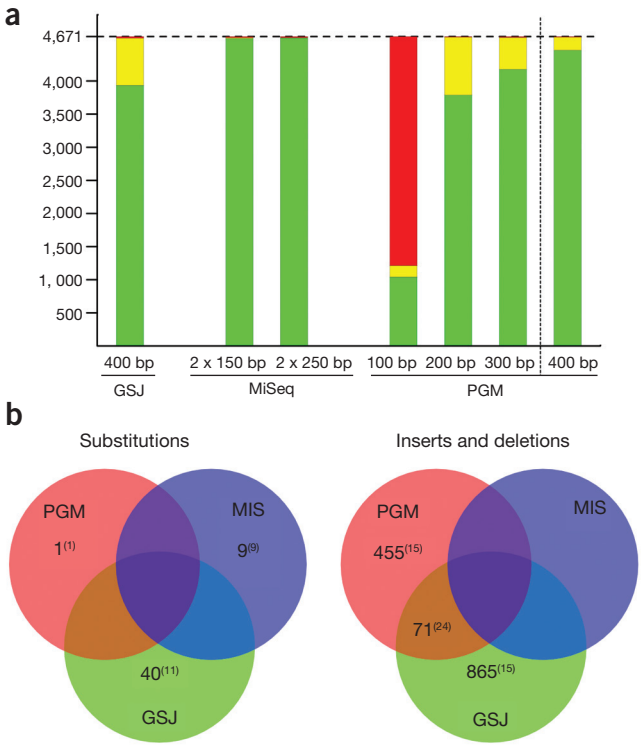

Figure 1 Evaluation of contiguity and consensus accuracy of draft *de novo* assemblies from benchtop sequencers. (**a**) Evolution of genome contiguity for GSJ, MiSeq and PGM. The contiguity of the *de novo* assembly consensus sequences generated by MIRA was analyzed for 4,671 nonpseudo- or nonparalogous chromosomal coding *E. coli* Sakai NCBI reference sequence genes. This genome-wide gene-by-gene analysis was performed with the Ridom SeqSphere+ software. Green segments denote 'perfect' genes that had a match of 100% identity and 100% overlap with the reference gene; yellow denotes genes that had a match of ≥97% identity and ≥97% overlap with the reference gene (but <100% identity and <100% overlap, owing mainly to indels); red indicates genes that had no match or <97% identity and/or <97% overlap with the reference gene counterpart. The 400-bp PGM chemistry was not officially available during time of study. (**b**) Venn diagram of consensus sequencing accuracy for PGM 300 bp, MiSeq 2 × 250-bp PE (MIS) and GSJ. Reported consensus errors were analyzed for 4,632 coding NCBI Sakai reference genome genes that could be retrieved from the MIRA *de novo* assemblies using SeqSphere+ for all three platforms. Numbers of variants confirmed by bidirectional Sanger sequencing are indicated in parentheses.

lengths (**Supplementary Table 2**). However, assemblies obtained from MiSeq 2 × 250-bp PE and PGM 300-bp data both had N50 sizes longer than assemblies from GSJ data.

Contiguity and consensus accuracy of draft *de novo* assemblies (not the raw read accuracy, read length or N50 values) are the most relevant parameters for the use of NGS in public-health microbiology. Contiguity describes the length of an assembled sequence contig and whether there are gaps or unresolved ambiguities[5]. Therefore, we performed a genome-wide, gene-by-gene analysis by examining the contiguity of the consensus assembled sequences for 4,671 nonpseudo- and nonparalogous coding Sakai genes (**Fig. 1a**). We obtained the best contiguity with assemblies based on MiSeq 2 × 150-bp PE data (99.6% of assembled genes were identical to and exactly the same length as the Sakai reference). The MiSeq 2 × 250-bp chemistry did not increase the number of 'perfect' recovered genes (99.4%), that is, genes that had a match of 100% identity and 100% overlap with the reference gene. This is in contrast to the PGM, for which every read-length update substantially increased the number of recovered genes. When used with the 300-bp chemistry, the PGM exceeded the GSJ, with 89.4% perfect recovered genes, versus 84.2% for the GSJ.

To assess the consensus accuracy of the *de novo* assemblies, we analyzed 4,632 coding genes from the US National Center for Biotechnology Information (NCBI) Sakai reference genome that could be retrieved from MIRA-generated *de novo* assemblies using SeqSphere+ version 0.99.20 (Ridom; Münster, Germany) for GSJ, PGM 300-bp and MiSeq 2 × 250-bp PE data (**Fig. 1b**). We used bidirectional Sanger sequencing, similarly to a previous study[6], to resolve a set of 98 discrepancies between our assemblies and the reference. This set comprised the 23 discrepancies that were common to all three platforms, the single PGM substitution and the nine MiSeq substitutions, 11 randomly selected GSJ substitutions and in total 54 randomly selected GSJ and PGM indels. Validation of the eight substitution and 15 indel variants common to all three NGS platforms suggested that either the Sakai strain sequenced here underwent micro-evolutionary changes or the genome sequence deposited in 2001 contains sequencing errors.

Data on genome-wide change rates and adaptation during long-term *in vitro* growth of *E. coli* has started to accumulate only recently[7]. The remaining 75 Sanger sequencing–controlled discrepancies turned out to be benchtop NGS sequencing errors. For substitutions, the PGM and MiSeq data, therefore, contained only one and nine consensus sequencing errors, respectively. In contrast, the 40 substitutions in the GSJ data indicated a slightly higher consensus error rate. No indel consensus errors were observed in the MiSeq data, whereas 526 and 936 indels were observed in the PGM and GSJ data, respectively.

Conducting a perfect desktop NGS technology comparison is nearly impossible. However, some crucial points could be addressed here, specifically in contrast to the previous study[1]. First, a strain that is available from culture collections was used as reference. Second, DNA from the same preparation was used for all sequencing reactions. Third, all NGS machines were operated by experienced academic users, thus avoiding the direct involvement of vendors of NGS machines. Neither we nor Loman *et al.*[1] evaluated benchtop sequencers over the whole range of bacterial GC content. The Wellcome Trust Sanger Institute (Cambridge, UK) recently studied this topic[8]. Unfortunately, the authors did not supply a repository with raw sequence data along with all applied software parameters and scripts. Any technology evaluation is a snapshot in time; for this study, we were able to use a 400-base template-preparation kit before it became publicly available at the end of 2012 (our PGM 400-bp beta-testing results (**Fig. 1a**, **Table 1** and **Supplementary Tables 1** and **2**) demonstrated a further increase in read length and 95.6% 'perfect', recovered Sakai genes when testing for contiguity).

Does our analysis throw light on which instrument a buyer should consider? As usual, no single platform comes out on top. The 454 GSJ wins when read length and number of established library protocols are taken into consideration. The Illumina MiSeq is best with respect to throughput per run and least number of consensus errors. Both the Ion Torrent PGM and the GSJ are well suited for sequencing amplicons, with the former showing greater recent improvements to the technology and throughput per day.

Accession codes. Sequencing reads produced during this study have been deposited at the NCBI Short Read Archive

under the accession number SRS352585. Assemblies, mapping files, analysis scripts and documentation have been uploaded to a public Github repository and are available at https://github.com/ngscomparison/NGS-Benchtop-Comparison.

Note: Supplementary information is available in the in the online version of the paper (doi:10.1038/nbt.2522).

ACKNOWLEDGMENTS

This work was supported by grants from the German Federal Ministry of Education and Research (grant no. 01KI1012B, FBI-Zoo project) and the European Commission's Seventh Framework Programme (EU PathoNGenTrace project grant agreement no. 278864) to D.H. and A.M. and from the German Research Foundation Deep Metazoan Phylogeny priority program SPP 1174 (grant no. HA1628/9-2) to A.v.H. U.J. was funded by the PACES research program of the Alfred Wegener Institute, within the Helmholtz Foundation Initiative in Earth and Environment.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper (doi:10.1038/nbt.2522).

Sebastian Jünemann[1,2], Fritz Joachim Sedlazeck[3], Karola Prior[1], Andreas Albersmeier[4], Uwe John[5], Jörn Kalinowski[4], Alexander Mellmann[6], Alexander Goesmann[2,7], Arndt von Haeseler[3], Jens Stoye[2,8] & Dag Harmsen[1]

[1]Department for Periodontology, University of Münster, Münster, Germany. [2]Institute for Bioinformatics, Center for Biotechnology, Bielefeld University, Bielefeld, Germany. [3]Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna and Medical University of Vienna, Vienna, Austria. [4]Technology Platform Genomics, Institute for Genomic Research and Systems Biology, Center for Biotechnology, Bielefeld University, Bielefeld, Germany. [5]Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany. [6]Institute of Hygiene, University of Münster, Münster, Germany. [7]Bioinformatics Resource Facility, Center for Biotechnology, Bielefeld University, Bielefeld, Germany. [8]Genome Informatics, Faculty of Technology, Bielefeld University, Bielefeld, Germany.

e-mail: dahrmsen@uni-muenster.de

1. Loman, N.J. *et al. Nat. Biotechnol.* **30**, 434–439 (2012).
2. Hayashi, T. *et al. DNA Res.* **8**, 11–22 (2001).
3. Kingsford, C., Schatz, M. & Pop, M. *BMC Bioinformatics* **11**, 21 (2010).
4. Anonymous. *De novo* assembly using Illumina reads. (Illumina, 2010) <http://www.illumina.com/Documents/products/technotes/technote_denovo_assembly_ecoli.pdf>.
5. Felsenfeld, A. *et al. Genome Res.* **9**, 1–4 (1999).
6. Lam, H.Y.K. *et al. Nat. Biotechnol.* **30**, 78–82 (2012).
7. Barrick, J.E. *et al. Nature* **461**, 1243–1247 (2009).
8. Quail, M.A. *et al. BMC Genomics* **13**, 341 (2012).

Loman *et al.* reply:

We were pleased to see this useful update from Jünemann *et al.*[1] to our article 'Performance comparison of benchtop sequencers'[2]. Progress in sequencing technologies is driving genomic research at an astonishing rate. More than 14 months have elapsed since we submitted our manuscript based on data generated in the summer of 2011. There have been impressive changes in throughput (up to fivefold) and read length (up to fourfold) during this time, easily outperforming Moore's Law. However, we note that despite these improvements, our overall conclusions on the relative performance of the 454 GS Junior, Ion Torrent Personal Genome Machine (PGM) and Illumina MiSeq benchtop sequencers remain unchanged.

One anomalous issue in this article is the large discrepancy between the reported insertion and deletion (indel) rates from our two runs, of 316 chips, in July 2011 and those reported by Jünemann *et al.*[1]. Without access to the data, we can only speculate about the reason, but it seems probable that the discrepancy is related to the different read-trimming procedures used. More stringent read-trimming algorithms are likely to result in an improvement in error rate, as there is a strong correlation between quality score and actual error rate (as noted in our original study)[2]. We note that other, contemporaneous studies describe error rates for the PGM equivalent to those that we reported: in one study[3] the total error rate was 1.78%, and in a second study[4] an insertion rate of 0.693% and deletion rate of 0.965% were reported. We also note that the 100-base-pair data set generated by Jünemann *et al.*[1] on the 316 chip, contemporaneously with our study, performed particularly badly during *de novo* assembly with an N50 <1.5 kb and did not allow the vast majority of coding sequences in the *Escherichia coli* Sakai genome to be reconstructed without errors. Such poor assembly statistics at high coverage are hard to reconcile with the low error rates quoted by Jünemann *et al.*[1].

There is no sign that progress in genome sequencing technologies is slowing. Publication delays have the potential to limit the use of such platform comparisons, but we believe these comparisons are nonetheless more useful than marketing literature or anecdotes. We would welcome a community-led, open-access project to provide trustworthy benchmarking in a timely and objective fashion.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Mark John Pallen

University of Birmingham, Birmingham, UK.
email: m.pallen@bham.ac.uk

1. Jünemann, S. *et al. Nat. Biotechnol.* **31**, 294–295 (2013).
2. Loman, N.J. *et al. Nat. Biotechnol.* **30**, 434–439 (2012).
3. Quail, M.A. *et al. BMC Genomics* **13**, 341 (2012).
4. Liu, L. *et al. J. Biomed. Biotechnol.* **2012**, 251364 (2012).