Forensic Science International: Genetics Supplement Series 2 (2009) 518–519

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/FSIGSS

Research article

Application of full mitochondrial genome sequencing using 454 GS FLX pyrosequencing

Martin Mikkelsen [a,*], Eszter Rockenbauer [a], Andrea Wächter [a], Liane Fendt [b], Bettina Zimmermann [b], Walther Parson [b], Sandra Abel Nielsen [c], Tom Gilbert [c], Eske Willerslev [c], Niels Morling [a]

[a] Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark
[b] Institute of Legal Medicine, Innsbruck Medical University, Austria
[c] Natural History Museum of Denmark, University of Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 14 September 2009
Accepted 16 September 2009

Keywords:
454 pyrosequencing
Whole mtDNA genome sequencing

ABSTRACT

The GS FLX pyrosequencing platform using parallel tagged sequencing was tested on 10 Somali individuals for sequencing of the complete mitochondrial genome. The amplicons were sequenced twice with increasing coverage to establish the minimum of coverage needed to produce reliable sequence reads. The genome sequences were compared to previously obtained control regions sequences with Sanger sequencing and 49 SNPs in coding regions of the mitochondrial genome. No discrepancy was found with the three methods except in a poly-C stretch that was estimated to be 16193.1C by sequencing with terminator chemistry and no insert was observed with pyrosequencing. The sequence lengths of both pyrosequencing runs were between 225 and 235 nucleotides. The coverage for the first pyrosequencing run was between 0 and 159 and 10–577 for the second one.

1. Introduction

Sequencing of the HV1 and HV2 mitochondrial regions has been a valuable tool in forensic and population genetics for many years. However, it is of highest interest to determine the sequence of all the variable sites throughout the complete mitochondrial genome. Complete genome sequencing of the mitochondrial genome is now possible in a cost and time efficient manner using next generation DNA sequencing platforms. The aim of this study was to test the performance of parallel DNA sequencing with the GS FLX platform from 454 Life Sciences.

2. Materials and methods

The complete mitochondrial genome from each of 10 unrelated Somali individuals was amplified in two parallel PCRs creating overlapping amplicons. The PCR amplicons were pooled in equimolar ratios. Parallel tagged sequencing was employed to allow pyrosequencing of several samples in parallel [1].

The GS FLX platform was used to establish the template library for pyrosequencing. The samples were sequenced twice. The first pyrosequencing run was on 1/8 of a sequencing plate together with 10 other samples (data not shown). The second run was performed on 1/4 of a plate with only the 10 Somali samples.

The sequences were sorted according to their individual tags, which were then removed. The sequences were assembled around the rCRS [2] using the software Geneious 4.7.4 (Biomatters Ltd., New Zealand). The sequences were inspected for obvious alignments errors, misreads, single base insertions and deletions, chimera sequence artefacts and remaining tag sequences. A consensus sequence was generated using the majority setting in Geneious. From the consensus sequence, a mutation report was generated using the software Sequencer 4.8 (GeneCodes, USA).

The sequences obtained with pyrosequencing were compared to partial sequences generated with Sanger sequencing of position 16,024 to 576 in the control region or the nucleotides of 49 SNPs in coding regions typed by means of single base extension with the SNaPshot kit (Applied Biosystems, USA).

3. Results and discussion

The sequence lengths of the two runs were 225–235 nucleotides. The numbers of sequences per sample varied between 431 and 2375 in the first run, and between 2489 and 11,704 in the second run (Table 1). Variation in the coverage was observed along the genome and between the samples. High coverage was observed in the two areas where the initial PCR primers overlap. The mean coverage in the first pyrosequencing run was 16.20, range: 0–159 (Table 1). In the second run, the mean coverage was 99.73, range:

* Corresponding author. Tel.: +45 35326287; fax: +45 35326270.
E-mail address: martin.mikkelsen@forensic.ku.dk (M. Mikkelsen).

1875-1768/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigss.2009.09.028

Table 1
Coverage, sequences and reads generated from the two sequencing runs.

	1st run					2nd run				
	Coverage			No. of sequences	No. of base reads	Coverage			No. of sequences	No. of base reads
	Min.	Max.	Mean	#	Mb	Min.	Max.	Mean	#	Mb
21	1	91	14.59	987	0.24	40	516	120.14	8,781	1.99
22	2	58	16.68	1127	0.28	62	472	150.35	10,939	2.49
23	1	54	14.80	1006	0.25	43	438	122.46	9,014	2.03
24	0	31	6.29	431	0.10	10	148	34.86	2,489	0.57
25	2	62	16.27	1178	0.27	32	281	92.58	6,837	1.53
26	1	69	15.55	1117	0.26	42	339	89.36	6,493	1.48
27	0	50	9.90	716	0.16	16	158	51.15	3,774	0.85
28	2	78	14.33	1014	0.24	35	229	74.65	5,438	1.24
29	7	159	32.77	2375	0.54	45	577	157.70	11,704	2.61
30	3	76	20.80	1499	0.34	38	372	104.03	7,647	1.72

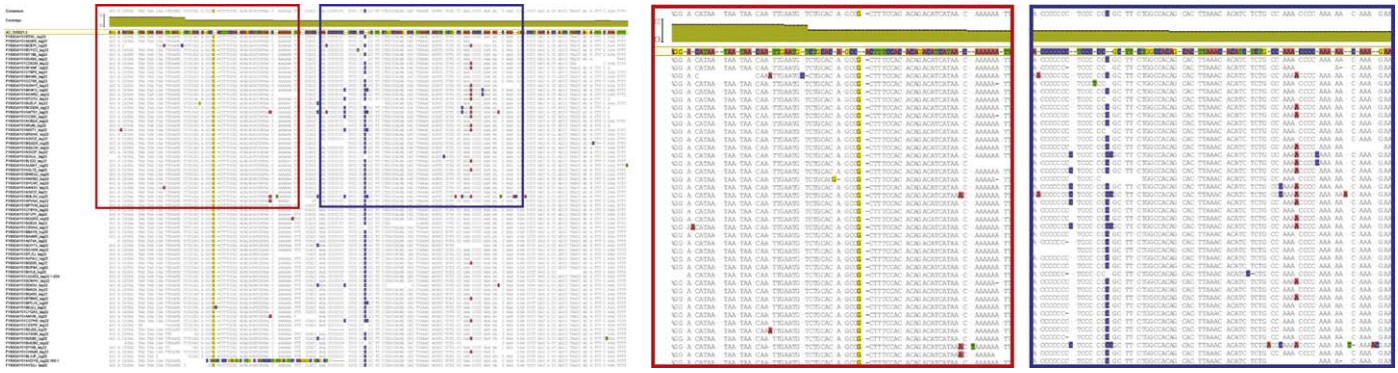

Fig. 1. Part of an mtDNA sequence assembly around the poly-C stretch at the location 315. The green bar indicates the coverage of the specific area. The top sequence above the green coverage graph is the generated consensus sequence. Below the green bar, the various sequences are listed with the reference sequence (rCRS) at the top. Deviates from the rCRS are highlighted in colour. Artefacts are seen in the homopolymeric stretch (blue box). Two true polymorphisms can be seen. The 263G (highlighted in yellow in the consensus sequence, red box) and 315.1C (highlighted in blue in the consensus sequence, blue box).

10–577. The coverages in the first pyrosequencing run were not sufficient to establish valid consensus sequences, while the coverages in the second run were sufficient to establish valid consensus sequences. It seems as if 30 samples on 1/4 of a plate allowing for approximately 120 samples per run would be a cost effective set-up.

The comparison of the results of pyrosequencing with previously obtained data differed only in one sample in a stretch, in which an insertion (16193.1C) was called in the poly-C stretch in HV1 by Sanger sequencing. With pyrosequencing, 34 of 136 sequences were reported as 16193.1C, while 102 sequences did not show an insert. Electropherograms of Sanger sequencing documented that the sample had length heteroplasmy at position 16,193.

Even when no length heteroplasmy is present, pyrosequencing does not seem to be robust concerning stretches of homopolymers such as the poly-C stretches in HV1 and HV2 (Fig. 1, blue box). However, if the coverage is high enough, a few artefacts will not contribute significantly to the consensus sequence.

The costs of reagents per sequence of mitochondrial genome is approximately 65€ using the GS FLX Platform. If the GS Titanium platform were used, the price would be approximately 50€ because this platform can produce longer sequences (400–500 bp). Preparation of samples and sequencing of up to 120 mitochondrial genomes can be performed in approximately 7 days by one person.

4. Conclusion

Investigations of mtDNA with the new generation sequencing tool GS FLX from 454 Life Sciences offers whole mitochondrial genome sequencing. The coverage must be high to obtain reliable results and the analysis software needs improvements to speed up the process and to decrease the need for human intervention.

Conflict of interest

None.

References

[1] M. Meyer, U. Stenzel, M. Hofreiter, Parallel tagged sequencing on the 454 platform, Nat. Protoc. 3 (2) (2008) 267–278.
[2] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat. Genet. 23 (2) (1999) 147–1147.