

GENETICS

454 OR 0815 OR 4911

22.12.2006

In a recent [book chapter](#) we discussed new genotyping and sequencing technologies. Our concluding remarks haven't changed so much – it seems that realtime detection of single molecules is still not possible; micro electrophoresis based methods have already reached their limit while sequencing by hybridization has severe restrictions when it comes to de novo (or re-) sequencing of whole genomes. At least for research purpose I expect that whole genome re-sequencing will replace current SNP based disease mapping. So far, sequencing by synthesis seems to be one of the few HT methods that already works at that scale. The 454 platform consists of 3 consecutive steps:

1. DNA library preparation starts with genomic DNA (after fragmentation and adaptor ligation the single-stranded template DNA libraries are isolated and assessed (takes ~4 hours)
2. sstDNA is emulsificated, then amplified and recovered on beads before sequencing primer are being annealed (takes 1 day)
3. After washing, so called PicoTiterPlates are prepared and a process started that looks like a combination of pyrosequencing reaction, correct me if I am wrong, a pyrophosphate dependent enzyme cascade emitting light being recorded by a CCD camera that watches each of the ~200,000 holes (takes ~6 hours according to a recent paper in GenomXPress 2/06, figures at [454.com](#))

With an average read length of 100 bp and 200,000 fragments (resulting in 20 Mb) in 6 hours, the throughput is about 60fold compared to Sanger sequencing. The recent [Neanderthal](#) paper raises five arguments why the 454 sequencing platform is extremely well suited for analyses of bulk DNA extracted from ancient remains.

1. ... it circumvents bacterial cloning, in which the vast majority of initial template molecules are lost during transformation and establishment of clones.
2. ... because each molecule is amplified in isolation from other molecules it also precludes template competition, which frequently occurs when large numbers of different DNA fragments are amplified together.
3. ... its current read length of 100–200 nucleotides covers the average length of the DNA preserved in most fossils.
4. ... it generates hundreds of thousands of reads per run, which is crucial because the majority of the DNA recovered from fossils is generally not derived from the fossil species, but rather from organisms that have colonized the organism after its death.
5. ... because each sequenced product stems from just one original single-stranded template molecule of known orientation, the DNA strand from which the sequence is derived is known. This provides an advantage over traditional PCR from double-stranded templates, in which the template strand is not known, because the frequency of different nucleotide misincorporations can be deduced ... damage that affects different bases differently.

Except of the low read length most of these observations would benefit large scale resequencing projects in human individuals. My main point for starting ASAP resequencing projects: So far we have not achieved a dense resolution of the genome while deep resequencing project (for example at the CRP locus) got astonishing results. We do not even know what is going on in the “noncoding” regions. Finally deletions and CNVs have been largely neglected – another look at this question in the [EJHG](#).

The question remains, what does 454 mean – my inquiry is still pending. As far as I know 0815 was a machine gun in World War I and is a synonym for something repeatedly boring while 4911 is a street number and stands for a perfume. Yea, yea.