Sequencing and de novo assembly of 150 genomes from Denmark as a population reference

Sequencing and de novo assembly of 150 genomes from Denmark as a population reference

Hundreds of thousands of human genomes are now being sequenced to characterize genetic variation and use this information to augment association mapping studies of complex disorders and other phenotypic traits. Genetic variation is identified mainly by mapping short reads to the reference genome or by performing local assembly. However, these approaches are biased against discovery of structural variants and variation in the more complex parts of the genome. Hence, large-scale de novo assembly is needed. Here we show that it is possible to construct excellent de novo assemblies from high-coverage sequencing with mate-pair libraries extending up to 20 kilobases. We report de novo assemblies of 150 individuals (50 trios) from the GenomeDenmark project. The quality of these assemblies is similar to those obtained using the more expensive long-read technology. We use the assemblies to identify a rich set of structural variants including many novel insertions and demonstrate how this variant catalogue enables further deciphering of known association mapping signals. We leverage the assemblies to provide 100 completely resolved major histocompatibility complex haplotypes and to resolve major parts of the Y chromosome. Our study provides a regional reference genome that we expect will improve the power of future association mapping studies and hence pave the way for precision medicine initiatives, which now are being launched in many countries including Denmark.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Metagenomics, Department of Systems Biology, Center for Biological Sequence Analysis, Integrative Systems Biology, Integrative Systems Biology, Disease Intelligence and Molecular Evolution, Genomic Epidemiology, High Performance Computing, Functional Human Variation, University of Copenhagen, Aarhus University, BGI-Shenzhen , BGI Europe A/S, Technical University of Denmark, University of Oslo, University of Bergen, Karolinska Institutet
Pages: 87-91
Publication date: 2017
Peer-reviewed: Yes

Publication information
Journal: Nature
Volume: 548
ISSN (Print): 0028-0836
Ratings:
BFI (2018): BFI-level 3
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): CiteScore 14.59
Web of Science (2017): Impact factor 19.181
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 13.33
Web of Science (2016): Impact factor 19.304
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): CiteScore 14.38
Web of Science (2015): Impact factor 17.184
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 14.22
Web of Science (2014): Impact factor 14.547
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): CiteScore 14.96