Analysis of 62 hybrid assembled human Y chromosomes exposes rapid structural changes and high rates of gene conversion

The human Y-chromosome does not recombine across its male-specific part and is therefore an excellent marker of human migrations. It also plays an important role in male fertility. However, its evolution is difficult to fully understand because of repetitive sequences, inverted repeats and the potentially large role of gene conversion. Here we perform an evolutionary analysis of 62 Y-chromosomes of Danish descent sequenced using a wide range of library insert sizes and high coverage, thus allowing large regions of these chromosomes to be well assembled. These include 17 father-son pairs, which we use to validate variation calling. Using a recent method that can integrate variants based on both mapping and de novo assembly, we genotype 10898 SNVs and 2903 indels (max length of 27241 bp) in our sample and show by father-son concordance and experimental validation that the non-recurrent SNP and indel variation on the Y chromosome tree is called very accurately. This includes variation called in a 0.9 Mb centromeric heterochromatic region, which is by far the most variable in the Y chromosome. Among the variation is also longer sequence-stretches not present in the reference genome but shared with the chimpanzee Y chromosome. We analyzed 2.7 Mb of large inverted repeats (palindromes) for variation patterns among the two palindrome arms and identified 603 mutation and 416 gene conversions events. We find clear evidence for GC-biased gene conversion in the palindromes (and a balancing AT mutation bias), but irrespective of this, also a strong bias towards gene conversion towards the ancestral state, suggesting that palindromic gene conversion may alleviate Muller's ratchet. Finally, we also find a large number of large-scale gene duplications and deletions in the palindromic regions (at least 24) and find that such events can consist of complex combinations of simultaneous insertions and deletions of long stretches of the Y chromosome.

General information

State: Published
Organisations: Department of Bio and Health Informatics, Integrative Systems Biology, Department of Systems Biology, Center for Biological Sequence Analysis, Integrative Systems Biology, Disease Intelligence and Molecular Evolution, Genomic Epidemiology, Functional Human Variation, Aarhus University, Technical University of Denmark, University of Bergen, Karolinska Institutet, BGI Europe A/S, University of Bristol, University of Copenhagen, BGI-Shenzhen
Number of pages: 20
Publication date: 2017
Peer-reviewed: Yes

Publication information
Journal: P L o S Genetics
Volume: 13
Issue number: 8
Article number: e1006834
ISSN (Print): 1553-7390
Ratings:
BFI (2019): BFI-level 2
Web of Science (2019): Indexed yes
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): CiteScore 5.57 SJR 4.829 SNIP 1.364
Web of Science (2017): Impact factor 5.54
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 5.93 SJR 5.457 SNIP 1.512
Web of Science (2016): Impact factor 6.1
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2