A benchmarking of human Y-chromosomal haplogroup classifiers from whole-genome and whole-exome sequence data

Víctor García-Olivares[1,2], Adrián Muñoz-Barrera[1], Luis A. Rubio-Rodríguez[1], David Jáspez[1], Ana Díaz-de Usera[1], Antonio Iñigo Campos[1], Krishna R. Veeramah[3], Santos Alonso[4,5], Mark G. Thomas[6,7], José M. Lorenzo-Salazar[1], Rafaela González-Montelongo[1,2], Carlos Flores[1,2,8,9,10*]

1. Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Spain.
2. Plataforma Genómica de Alto Rendimiento para el Estudio de la Biodiversidad, Instituto de Productos Naturales y Agrobiología (IPNA), Consejo Superior de Investigaciones Científicas, San Cristóbal de La Laguna, Spain.
3. Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, United States.
4. Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country UPV/EHU, Leioa, Bizkaia, Spain.
5. Biotechnology Center, Human Molecular Evolution Lab 2.08 UPV/EHU Science Park. 48940 Leioa, Bizkaia. Spain.
6. UCL Genetics Institute, University College London (UCL), Gower Street, London WC1E 6BT, United Kingdom.
7. Research Department of Genetics, Evolution & Environment, University College London (UCL), Darwin Building, Gower Street, London WC1E 6BT, United Kingdom.
8. Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain.
9. CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain.
10. Facultad de Ciencias de la Salud, Universidad Fernando de Pessoa Canarias, Las Palmas de Gran Canaria, Spain.

*Corresponding author: cflores@ull.edu.es

Abstract

The non-recombinant region of the Y chromosome (NRY) contains a great number of polymorphic markers that allows to accurately reconstruct pedigree relationships and retrieve ancestral information from study samples. The analysis of NRY is typically implemented in anthropological, medical, and forensic studies. High-throughput sequencing (HTS) has profoundly increased the identification of genetic markers in the NRY genealogy and has prompted the development of automated NRY haplogroup classification tools. Here, we present a benchmarking study of five command-line tools for NRY haplogroup classification. The evaluation was done using empirical short-read HTS data from 50 unrelated donors using paired data from whole-genome sequencing (WGS) and whole-exome sequencing (WES) experiments. Besides, we evaluate the performance of the top-ranked tool in the classification of data of third generation HTS obtained from a subset of donors. Our findings demonstrate that WES can be an efficient approach to infer the NRY haplogroup, albeit generally providing a lower level of genealogical resolution than that recovered by WGS. Among the tools evaluated, YLeaf offers the best performance for both WGS and WES applications. Finally, we demonstrate that YLeaf is able to correctly classify all samples sequenced with nanopore technology from long noisy reads.

52 ## 1. Introduction

53 The Y chromosome (chrY) is one of the smallest chromosomes in the human genome (~60 Mb). A high
54 proportion of this chromosome (95%), known as the non-recombining region (NRY), shows a patrilineal
55 inheritance following a haploid behavior due to its lack of chromosomal recombination during meiosis
56 (Quintana-Murci & Fellous, 2001). Because of this, the NRY allows for a precise reconstruction of its
57 genealogy back to a common ancestor, as described by the coalescence theory (Underhill & Kivisild,
58 2007). The study of NRY has a wide range of applications in fields such as evolutionary anthropology
59 and population history (Pinotti et al., 2019; Zeng et al., 2018), medical genetics (Colaco & Modi, 2018;
60 Grassmann et al., 2019), and forensic science (Kayser, 2017; Zhou et al., 2022).

61 The advent of high-throughput sequencing (HTS) technology has brought about a revolution in
62 the development of human genomics and medicine. The decrease in costs and the increase in coverage
63 of both whole-genome sequencing (WGS) and whole-exome sequencing (WES) applications offer the
64 possibility of improving the chrY research through deeper and more optimal analyses (Anderson et al.,
65 2019; Levy & Myers, 2016). However, this chromosome presents regions that are challenging to
66 sequence, such as short tandem repeats (STRs) (Alvarez-Cubero et al., 2018; Charlesworth, 2003). In
67 this regard, third-generation HTS with the development of longer reads with greater read depth may
68 help to improve the mapping of such complex repeat sequences (Anderson et al., 2019; Jobling & Tyler-
69 Smith, 2017). An example of this is the use of nanopore technology (ONT, Oxford Nanopore
70 Technologies, Oxford, UK) to successfully generate the first African human chrY reference assembly
71 (Kuderna et al., 2019).

72 The use of HTS technology for human genome sequencing has enabled the discovery of new
73 variants in the chrY providing a remarkable increase in the volume of marker information available to
74 trace the human paternal lineages (Claerhout et al., 2021). In this regard, the International Society of
75 Genetic Genealogy (ISOGG-Y-DNA tree; https://isogg.org/tree/) has compiled since 2006 all new
76 variants in the NRY, generating a database that currently hosts more than 90,650 unique biallelic
77 variants. The NRY diversity has been structured following a phylogenetic hierarchy based on variants
78 that define distinct clades representing haplotypes commonly referred to as haplogroups (The Y
79 Chromosome Consortium, 2002). The study of these haplogroups allows to trace origins, patterns of
80 differentiation between populations, and to unravel historical patterns of human migration over time

81 (Calafell & Larmuseau, 2017). Haplogroup identification is, therefore, a key step in recovering pedigree
82 relationships and ancestral information from analyzed samples.
83 The exponential increase in the number of NRY markers, which concomitantly associates with
84 a rise in the complexity of the chrY tree, and the development of HTS imposes a bioinformatics
85 challenge for inferring the patrilineal genetic genealogy of the study samples. To take advantage of the
86 potential offered by HTS technology, the number of automated NRY classification tools has seen a
87 considerable increase in recent years (Chen et al., 2021; David Poznik, 2016; Jagadeesan et al., 2020;
88 Martiniano et al., 2022; Ralf et al., 2018, 2019; Van Geystelen et al., 2013; Zhang et al., 2013). However,
89 an unmet need to date is a comparative study evaluating the performance of each tool. Here we present
90 a benchmarking analysis of several command-line tools for automated human NRY classification using
91 empirical short-read HTS data from two of the most widely used applications in human genetics, WGS
92 and WES. In addition, we assessed the performance of the best performing tool in third-generation HTS
93 long noisy read WGS data obtained by nanopore technology for a subset of the donors.
94

95 ## 2. Materials and Methods

96 ***Samples, library preparation, and sequencing***

97 The study was approved by the Research Ethics Committee of the Hospital Universitario Nuestra
98 Señora de Candelaria (CHUNSC_2020_95) and performed according to The Code of Ethics of the
99 World Medical Association (Declaration of Helsinki).
100 Fifty DNA samples from unrelated donors were used for the study after informed consent. All
101 samples were sequenced in parallel using short-read WGS and WES. The construction of libraries was
102 performed with Illumina preparation kits (**Table S1**) following the manufacturer's recommendations
103 (Illumina Inc., San Diego, CA, USA). The Nextera DNA Prep kit and Illumina DNA Prep were used for
104 WGS. The same samples were processed with Nextera DNA Exome and Illumina DNA Prep with
105 Enrichment as described elsewhere (Díaz-de Usera et al., 2020). The library quality controls were
106 carried out in a TapeStation 4200 (Agilent Technologies, Santa Clara, CA, USA) and sequencing was
107 conducted on HiSeq 4000 or NovaSeq 6000 (Illumina, Inc., San Diego, CA, USA) instruments.
108 Seven of these samples were also sequenced using long noisy read WGS data obtained with
109 nanopore technology at KeyGene (Wageningen, The Netherlands). Sequencing was performed on a
110 PromethION system (ONT) for 64 h using one FLO_PR002 (R9.4.1 pore) flow cell per sample following

111 the manufacturer's recommendations. Basecalling was conducted on the PromethION computing
112 module using MinKNOW v1.14.2 with Guppy v2.2.2, and data preprocessing metrics were carried out
113 with PycoQC v2.5.2 (Leger & Leonardi, 2019).
114
115 ***Bioinformatic processing***
116 Processing of short-read WGS and WES data was carried out using an in-house pipeline based on
117 GATK v4.1 for WGS and GATK v3.8 for WES (McKenna et al., 2010). Raw reads were assessed using
118 FastQC v.0.11.8 software (Andrews, 2010) and aligned to the GRCh37/hg19 reference genome using
119 BWA-MEM v0.7.15 (Li & Durbin, 2009). Quality control of aligned reads was performed with Qualimap
120 v2.2.1 (García-Alcalde et al., 2012). The alignments were then processed for duplicate marking and
121 base quality score recalibration (DePristo et al., 2011). Variant calling step was conducted by GATK
122 HaplotypeCaller following the Broad Institute's best practices workflow for germline short variant
123 discovery. From the resulting BAM files, the NRY region (2.64-59.03 Mb) was extracted by using
124 SAMtools v1.12 (Li et al., 2009). Regarding the WES data, it was observed that of the 676 DNA capture
125 probes used for hybridisation-based target enrichment of the chrY regions, 90% are within the NRY,
126 covering 0.22% of this region. For ONT data, raw long noisy reads were first preprocessed with FiltLong
127 v0.2.1 (https://github.com/rrwick/Filtlong) to filter reads shorter than 1,000 bp. Filtered reads were
128 assessed using NanoPlot v1.38.1 (De Coster et al., 2018), then aligned to the GRCh37/hg19 reference
129 genome using Minimap2 v2.22-r1101 (Li, 2018) and sorted with SAMtools v1.14, extracting only reads
130 aligned to the NRY region. All these bioinformatic processes were computed at TeideHPC infrastructure
131 (https://teidehpc.iter.es/en/home/).
132
133 ***Sex quality control***
134 To identify the sex of the donnors, a quality control was performed based on both the self-reported sex
135 of the individual and two bioinformatics approaches. The first approach, performed by Somalier v0.2.15
136 (Pedersen et al., 2019), identifies the sex of the sample from the depth of the X and Y chromosome
137 reads. For the second approach, an in-house heuristic script (https://github.com/genomicsITER/sexQC-
138 for-NGS-data) was used. The analysis involves assessing the depth of 11 genes in the non-
139 pseudoautosomal regions of the X and Y chromosomes using high quality mapped reads

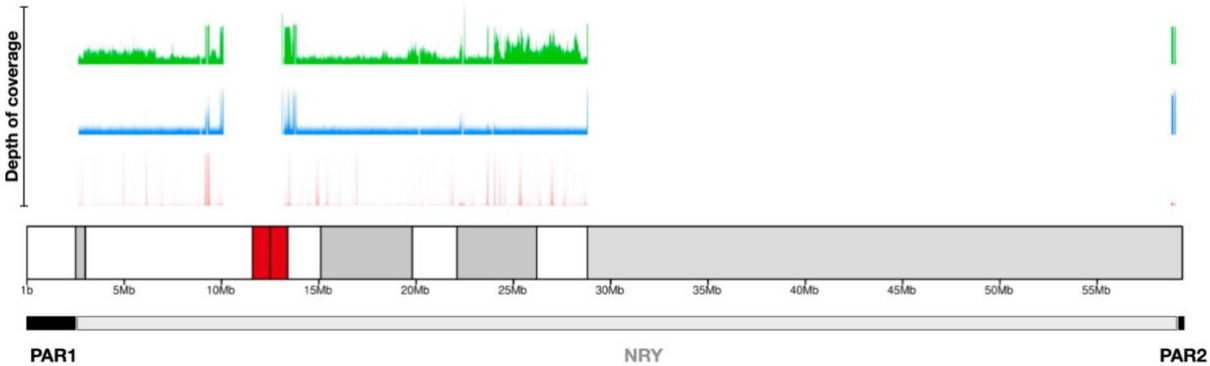
140 (MappingQuality>50). All samples used for this study were identified as males by consensus from these
141 two approaches.
142
143 ***Y-chromosomal haplogroup classification***
144 Among the tools available from the literature, we selected eight tools that were open-source and offered
145 a command-line interface (**Table 1**). These were run with the 2020 version (v15.73) of the ISOGG
146 repository database, which contains more than 90,000 polymorphic markers and it constitutes the
147 central reference for many bioinformatic tools to classify human chrY sequences. However, three of the
148 tools (YHap, AMY-tree, and yhaplo) were finally excluded from the study because they imposed
149 limitations for database updates. For YLeaf, the version 2.2 was used since the newer 3.1 version does
150 not use the ISOGG marker identifiers in the classification. Y-Lineage Tracker has the option of using
151 VCF and BAM input files, fostering an evaluation with the two alternative supported files. The
152 haplogroup classification process was executed using a workstation running CentOS 7 with 2 Intel Xeon
153 Cascade Lake 6252 Gold CPUs at 2.1 GHz and 384 GB of RAM. All tools were run using the default
154 parameters.
155 Unlike WGS, which recovers a larger part of the NRY, WES only partially recovers the NRY
156 (**Table S1**). This difference may lead to discrepancies in the haplogroup classification obtained by both
157 applications simply because it is expected that a lower level of resolution could be obtained for WES in
158 any given sample. To deal with this limitation in the benchmarking, we took as the reference for the
159 comparisons the maximum classification level retrieved by WES that matches the one obtained from
160 the WGS data.
161 **Table 1.** List of tools assessed for human Y-chromosomal haplogroup classification. All the tools assessed classify according to
162 the ISOGG nomenclature by using the latest version 15.73 (2020).
163

Tool	Release year	Version	ISOGG version (year)	Input options
pathPhynder	2022	1.a	2020	BAM
Y-LineageTracker	2021	1.3.0	2019	BAM/VCF
haploGrouper	2020	-	2019	VCF
clean_tree_v2	2019	2.0	2018	BAM
Yleaf	2018	2.2	2019	BAM
yhaplo	2016	1.1.0	2016	VCF
YHap	2013	-	2017	VCF
AMY-tree	2013	2.0	2013	VCF

164 ## 3. Results

165 ***Sequencing summary for short-read and long-read sequencing***

166 The mean (± SD) number of NRY reads recovered per sample (n = 50) for short-read WGS and WES
167 data were 8,329,867 ± 2,460,724 and 575,090 ± 176,201, respectively (**Table S1**). For WGS, 33.66%
168 of the NRY was covered at least at 10X. For WES, this percentage decreased to 0.9%. However, if only
169 the exonic regions of NRY are taken into account, 84.46% was covered by WES at least at 10X. The
170 mean (± SD) depth of coverage recovered across the NRY region for WGS was 13X ± 4 (range: 6-28X).
171 The depth decreased to less than 1X for WES, although it was as high as 67X ± 19 (range: 27-111X)
172 when only the exonic regions were included in the analysis. As for the detected single nucleotide
173 variants (SNVs), the mean (± SD) depth of coverage per SNV call had a value of 60 ± 17 in WGS,
174 decreasing to 32 ± 15 in WES. For ONT, the mean (± SD) number of NRY reads per sample (n = 7)
175 recovered was 177,436 ± 40,374. The mean NRY depth of coverage was 11X ± 2 (range: 8-14X) and
176 31.57% of the NRY was covered at least at 10X (**Table S1**). Furthermore, while the WGS data from
177 both sequencing technologies provided a homogeneous depth of coverage profile across the NRY
178 region (except in regions adjacent to the centromere and the heterochromatic region because of their
179 complexity), the WES data showed a heterogeneous profile with enriched sites (peaks) associated to
180 the capture of exons that are embedded within undetected regions (**Figure 1**).



181 **Figure 1**. Plot of the depth of coverage for short-read and long-read sequencing in the non-recombining portion of the Y
182 chromosome (NRY) of an exemplar sample. Long-read WGS data is shown in green. Short-read WGS and WES data are colored
183 in blue and red, respectively. In the ideogram of the Y chromosome, the heterochromatic regions (positive C-band) and the
184 centromere are colored in gray and red, respectively. In the lowest panel, the pseudoautosomal regions (PAR1 and PAR2) and
185 the NRY are represented in black and gray, respectively. To harmonize the results obtained from the three approaches, the depth
186 of coverage was normalized to 100X. The R package karyotypeR v1.2.2 (Gel & Serra, 2017) was used to generate the depth of
187 coverage plot.
188
189
190

191 ***Consensus haplogroup classification***

192 Based on the metrics retrieved for the short-read data, WGS showed higher values than WES for both
193 the breadth and depth of coverage parameters, both of which are closely related to higher statistical
194 support for variant detection. Therefore, we establish the WGS-derived haplogroup as the ground truth.
195 To assign the haplogroup of each sample, the haplogroup most frequently classified by all the tools
196 assessed was used as consensus. Fourteen samples showed 100% concordance among the tools
197 evaluated, and the remaining, we obtained a mean concordance rate of 66.20% (**Table S2**). However,
198 in most cases where discordance was reported, they were due to distinct haplogroup level
199 classifications and not to misclassification. In four of the samples inconsistencies among the tools
200 precluded a straightforward indication of the consensus haplogroup. In these four cases, the
201 classification result of YLeaf was used as the ground truth given its higher classification accuracy
202 demonstrated in all other samples. Considering only the WGS results, YLeaf offered the highest
203 classification accuracy (94%). With slightly lower, but relatively high performance (>70%), we found
204 clean_tree_v2, LineageTracker (with VCF as the input file), Haplogrouper, and pathPhynder. The worst
205 performing tool was LineageTracker using BAM as the input file since it misclassified 56% of the
206 samples. Based on the limitations outlined in the methodology and in order to harmonize the results,
207 the consensus haplogroup per sample used for the benchmarking was subordinated to the maximum
208 level of resolution retrieved by WES (**Table S2**).

209

210 ***Haplogroup classification***

211 The classification accuracy provided by short-read WGS data reached an average of 88.67%, while for
212 WES it decreased to an average of 43.67% (**Table S2**). On average (± SD), there were less
213 discordances on the WGS classifications (0.68 ± 0.96) than for WES (3.38 ± 1.01). By the classification
214 accuracy for WGS data, LineageTracker (VCF as input file) and YLeaf showed the highest accuracy,
215 classifying precisely 98% and 96% of the analyzed samples, respectively. With slightly lower accuracy
216 were the following tools: pathPhynder (92%), Haplogrouper (90%), and clean_tree_v2 (88%).
217 LineageTracker, using BAM as input file, was the least accurate tool for WGS data, misclassifying 32%
218 of the samples. For WES data, YLeaf showed the highest classification accuracy among all tools,
219 classifying precisely 88% of the analyzed samples. With a slightly lower accuracy than YLeaf were
220 clean_tree_v2 and pathPhynder, with an average of 80%. Haplogrouper and LineageTracker (VCF as

221 input file) were the least accurate tools, yielding an incorrect haplogroup classification in more than 88%
222 of the samples. LineageTracker from a BAM file did not provide a result for any of the samples. Since
223 YLeaf showed the best performance for both WGS and WES short-read data, this tool was used to
224 classify samples from ONT data. The seven samples sequenced with long noisy ONT reads were
225 correctly classified at the haplogroup level retrieving the same classification resolution as the WGS
226 short-read data.
227

228 ## 4. Conclusions

229 The advent of HTS technologies, the notable increase in the number of NRY polymorphic markers
230 detected, together with the importance of recovering ancestral information and pedigree relationships
231 of study samples, have motivated the development of new automated classification tools to adapt to
232 these challenges. In this study, we present a benchmarking of five classification tools that could be
233 easily upgraded to new versions of the ISOGG-Y-DNA tree. The comparison was carried out with
234 empirical paired HTS data from WGS and WES, two of the most widely used applications in human
235 genetics and medicine. Our results support that WES provides sufficient information to classify the NRY
236 haplogroup. Besides, we demonstrate that YLeaf shows the best performance among all the tools
237 evaluated for both applications, although with a slight loss in classification accuracy on the WES data.
238 In most of the samples, the classification retrieved matched with that inferred by WGS, although in
239 several samples a relatively lower level of accuracy was observed. However, considering that WES
240 sequences include a limited fraction of the NRY region, the performance achieved by YLeaf tool for
241 WES data is remarkable. Regarding third-generation HTS, our findings show that, despite the lower
242 *per-base* accuracy currently offered by the assessed technology, it did not preclude an equally accurate
243 classification as that obtained from short-read data.

5. Acknowledgements

We would like to thank the support from our colleagues from the Teide-HPC Supercomputing facility (http://teidehpc.iter.es/en), which was funded by INP-2011-0063-PCT-430000-ACT (INNPLANTA program) from the Spanish Ministry of Economy and Competitiveness.

6. Author Contributions

Conceptualization, V.G.O., A.M.B., L.A.R.R., C.F.; Data curation, V.G.O., A.M.B., L.A.R.R., A.M.B., J.M.L.S., D.J.T., R.G.M. K.R.V., M.G.T.; Methodology, V.G.O., A.M.B., J.M.L.S., L.A.R.R., A.D.U., D.J.T., A.I.C., R.G.M.; Supervision, C.F.; Writing—original draft preparation, V.G.O., A.M.B., C.F.; Writing—review and editing, V.G.O., A.M.B., L.A.R.R., S.A., C.F. All authors have read and agreed to the published version of the manuscript.

7. Conflicts of Interest

The authors declare no potential conflicts of interest with respect to the authorship and/or publication of this article. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

8. Funding

This research was funded by Ministerio de Ciencia e Innovación (RTC-2017-6471-1; AEI/FEDER, UE), co-financed by the European Regional Development Funds ‘A way of making Europe’ from the European Union; Ministerio de Economía, Industria y Competitividad (CGL2017-89021-P, Agencia Estatal de Investigación [AEI]), co-financed by European Regional Development Funds ‘A way of making Europe’ from the European Union; Fundación CajaCanarias and Fundación Bancaria “La Caixa” (2018PATRI20); Cabildo Insular de Tenerife (CGIEU0000219140); and by the agreement OA17/008 with Instituto Tecnológico y de Energías Renovables (ITER) to strengthen scientific and technological education, training, research, development and innovation in Genomics, Personalized Medicine and Biotechnology. ADU was supported by a fellowship from the Spanish Ministry of Education and Vocational Training (grant number FPU16/01435).

9. References

Alvarez-Cubero, M. J., Santiago, O., Martínez-Labarga, C., Martínez-García, B., Marrero-Díaz, R., Rubio-Roldan, A., Pérez-Gutiérrez, A. M., Carmona-Saez, P., Lorente, J. A., & Martinez-Gonzalez, L. J. (2018). Methodology for Y Chromosome Capture: A complete genome sequence of Y chromosome using flow cytometry, laser microdissection and magnetic streptavidin-beads. *Scientific Reports*, *8*(1), 9436.

Anderson, K., Cañadas-Garre, M., Chambers, R., Maxwell, A. P., & McKnight, A. J. (2019). The Challenges of Chromosome Y Analysis and the Implications for Chronic Kidney Disease. *Frontiers in Genetics*, *10*, 781.

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Calafell, F., & Larmuseau, M. H. D. (2017). The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research. *Human Genetics*, *136*(5), 559–573.

Charlesworth, B. (2003). The organization and evolution of the human Y chromosome. *Genome Biology*, *4*(9), 226.

Chen, H., Lu, Y., Lu, D., & Xu, S. (2021). Y-LineageTracker: a high-throughput analysis framework for Y-chromosomal next-generation sequencing data. In *BMC Bioinformatics* (Vol. 22, Issue 1). https://doi.org/10.1186/s12859-021-04057-z

Claerhout, S., Verstraete, P., Warnez, L., Vanpaemel, S., Larmuseau, M., & Decorte, R. (2021). CSYseq: The first Y-chromosome sequencing tool typing a large number of Y-SNPs and Y-STRs to unravel worldwide human population genetics. *PLoS Genetics*, *17*(9), e1009758.

Colaco, S., & Modi, D. (2018). Genetics of the human Y chromosome and its association with male infertility. *Reproductive Biology and Endocrinology: RB&E*, *16*(1), 14.

David Poznik, G. (2016). Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. In *Cold Spring Harbor Laboratory* (p. 088716). https://doi.org/10.1101/088716

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* , *34*(15), 2666–2669.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko,

A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498.

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T. F., & Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* , *28*(20), 2678–2679.

Gel, B., & Serra, E. (2017). karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. In *Bioinformatics* (Vol. 33, Issue 19, pp. 3088–3090). https://doi.org/10.1093/bioinformatics/btx346

Grassmann, F., Kiel, C., den Hollander, A. I., Weeks, D. E., Lotery, A., Cipriani, V., Weber, B. H. F., & International Age-related Macular Degeneration Genomics Consortium (IAMDGC). (2019). Y chromosome mosaicism is associated with age-related macular degeneration. *European Journal of Human Genetics: EJHG*, *27*(1), 36–41.

Jagadeesan, A., Ebenesersdóttir, S. S., Guðmundsdóttir, V. B., Thordardottir, E. L., Moore, K. H. S., & Helgason, A. (2020). HaploGrouper: A generalized approach to haplogroup classification. *Bioinformatics* . https://doi.org/10.1093/bioinformatics/btaa729

Jobling, M. A., & Tyler-Smith, C. (2017). Human Y-chromosome variation in the genome-sequencing era. *Nature Reviews. Genetics*, *18*(8), 485–497.

Kayser, M. (2017). Forensic use of Y-chromosome DNA: a general overview. *Human Genetics*, *136*(5), 621–635.

Kuderna, L. F. K., Lizano, E., Julià, E., Gomez-Garrido, J., Serres-Armero, A., Kuhlwilm, M., Alandes, R. A., Alvarez-Estape, M., Alioto, T., Gut, M., Gut, I., Schierup, M. H., Fornas, O., & Marques-Bonet, T. (2019). Selective single molecule sequencing and assembly of a human Y chromosome of African origin. *Nature Communications*. https://doi.org/10.1101/342667

Leger, A., & Leonardi, T. (2019). pycoQC, interactive quality control for Oxford Nanopore Sequencing. *Journal of Open Source Software*, *4*(34), 1236.

Levy, S. E., & Myers, R. M. (2016). Advancements in Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics*, *17*, 95–115.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* , *34*(18), 3094–3100.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* , *25*(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25*(16), 2078–2079.

Martiniano, R., De Sanctis, B., Hallast, P., & Durbin, R. (2022). Placing Ancient DNA Sequences into Reference Phylogenies. *Molecular Biology and Evolution*, *39*(2). https://doi.org/10.1093/molbev/msac017

Pedersen, B. S., Bhetariya, P. J., Brown, J., Marth, G., Jensen, R. L., Bronner, M. P., Underhill, H. R., & Quinlan, A. R. (2019). Somalier: rapid relatedness estimation for cancer and germline studies using efficient genome sketches. In *bioRxiv* (p. 839944). https://doi.org/10.1101/839944

Pinotti, T., Bergström, A., Geppert, M., Bawn, M., Ohasi, D., Shi, W., Lacerda, D. R., Solli, A., Norstedt, J., Reed, K., Dawtry, K., González-Andrade, F., Paz-Y-Miño, C., Revollo, S., Cuellar, C., Jota, M. S., Santos, J. E., Jr, Ayub, Q., Kivisild, T., … Tyler-Smith, C. (2019). Y Chromosome Sequences Reveal a Short Beringian Standstill, Rapid Expansion, and early Population structure of Native American Founders. *Current Biology: CB*, *29*(1), 149-157.e3.

Quintana-Murci, L., & Fellous, M. (2001). The Human Y Chromosome: The Biological Role of a "Functional Wasteland." *Journal of Biomedicine & Biotechnology*, *1*(1), 18–24.

Ralf, A., Montiel González, D., Zhong, K., & Kayser, M. (2018). Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data. *Molecular Biology and Evolution*, *35*(5), 1291–1294.

Ralf, A., van Oven, M., Montiel González, D., de Knijff, P., van der Beek, K., Wootton, S., Lagacé, R., & Kayser, M. (2019). Forensic Y-SNP analysis beyond SNaPshot: High-resolution Y-chromosomal haplogrouping from low quality and quantity DNA using Ion AmpliSeq and targeted massively parallel sequencing. *Forensic Science International. Genetics*, *41*, 93–106.

The Y Chromosome Consortium. (2002). A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups. In *Genome Research* (Vol. 12, Issue 2, pp. 339–348). https://doi.org/10.1101/gr.217602

Underhill, P. A., & Kivisild, T. (2007). Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annual Review of Genetics*, *41*, 539–564.

363 Van Geystelen, A., Decorte, R., & Larmuseau, M. H. D. (2013). AMY-tree: an algorithm to use whole
364 genome SNP calling for Y chromosomal phylogenetic applications. *BMC Genomics*, *14*, 101.
365 Zeng, T. C., Aw, A. J., & Feldman, M. W. (2018). Cultural hitchhiking and competition between patrilineal
366 kin groups explain the post-Neolithic Y-chromosome bottleneck. *Nature Communications*, *9*(1),
367 2077.
368 Zhang, F., Chen, R., Liu, D., Yao, X., Li, G., Jin, Y., Yu, C., Li, Y., & Coin, L. J. M. (2013). YHap: a
369 population model for probabilistic assignment of Y haplogroups from re-sequencing data. *BMC*
370 *Bioinformatics*, *14*, 331.
371 Zhou, Z., Zhou, Y., Li, Z., Yao, Y., Yang, Q., Qian, J., Shao, C., Qian, X., Sun, K., Tang, Q., & Xie, J.
372 (2022). Identification and assessment of a subset of Y-SNPs with recurrent mutation for forensic
373 purpose. *Forensic Science International*, *334*, 111270.
374

Supplementary Materials

Table S1. Short-read and long-read sequencing summaries of the samples.

Samples	*Whole-Genome Sequencing Illumina*					*Whole-Exome Sequencing Illumina*					*Whole-Genome Sequencing Oxford Nanopore*			
	Library Preparation	*Reads*	*Mean Depth of Coverage*	*Mean Depth per SNV*	*% NRY covered*	*Library Preparation*	*Reads*	*Mean Depth of Coverage*	*Mean Depth per SNV*	*% NRY covered*	*Library Preparation*	*Reads*	*Mean Depth of Coverage*	*% NRY covered*
ATTA022	Nextera-DNA-Prep	6,201,454	10	47	28	Illumina DNA Prep with Enrichment	307,731	49	32	69	-	-	-	-
BLA-050	Nextera-DNA-Prep	8,890,179	14	73	39	Illumina DNA Prep with Enrichment	534,379	96	58	91	-	-	-	-
CAM-GFF-117	Nextera-DNA-Prep	4,429,452	7	36	19	Illumina DNA Prep with Enrichment	370,722	69	34	79	-	-	-	-
CAM-PYKL-042	Nextera-DNA-Prep	10,896,378	17	76	40	Illumina DNA Prep with Enrichment	528,802	95	46	83	-	-	-	-
CAM-PYKL-083	Nextera-DNA-Prep	6,244,694	10	49	29	Illumina DNA Prep with Enrichment	508,906	82	38	77	-	-	-	-
CAM-PYKL-102	Nextera-DNA-Prep	5,116,461	8	42	21	Illumina DNA Prep with Enrichment	439,601	69	36	72	-	-	-	-
CAM-PYKL-140	Nextera-DNA-Prep	10,009,468	15	70	39	Illumina DNA Prep with Enrichment	601,355	111	49	84	-	-	-	-
CAM-PYKL-196	Nextera-DNA-Prep	6,953,783	12	55	37	Illumina DNA Prep with Enrichment	508,129	88	42	86	-	-	-	-
CANIS0005	Illumina-DNA-Prep	7,733,041	10	29	28	Nextera DNA Exome	602,741	54	21	86	-	-	-	-
CANIS0018	Illumina-DNA-Prep	6,861,802	9	59	24	Nextera DNA Exome	684,540	74	27	93	SQK LSK109 library prep	243,552	14	39
CANIS0025	Illumina-DNA-Prep	18712,303	28	10	40	Nextera DNA Exome	561,182	60	26	87	-	-	-	-
CANIS0031	Illumina-DNA-Prep	8,225,012	13	71	38	Nextera DNA Exome	721,281	64	28	93	-	-	-	-
CANIS0057	Nextera-DNA-Prep	10,390,095	17	69	40	Nextera DNA Exome	535,888	57	32	80	-	-	-	-
CANIS0897	Nextera-DNA-Prep	7,598,324	13	47	33	Nextera DNA Exome	373,294	44	21	76	-	-	-	-
CANIS0123	Illumina-DNA-Prep	4,094,921	6	38	13	Nextera DNA Exome	161,617	27	20	62	-	-	-	-
CANIS0140	Illumina-DNA-Prep	6,398,874	6	34	12	Nextera DNA Exome	917,472	99	26	94	SQK LSK109 library prep	155,743	10	28
CANIS0141	Illumina-DNA-Prep	7,600,080	13	60	36	Nextera DNA Exome	659,253	66	24	92	-	-	-	-
CANIS0166	Nextera-DNA-Prep	9,139,067	15	63	38	Nextera DNA Exome	642,086	57	29	87	-	-	-	-
CANIS0170	Illumina-DNA-Prep	9,297,218	14	76	36	Nextera DNA Exome	679,582	62	27	91	-	-	-	-
CANIS0248	Nextera-DNA-Prep	8,168,328	14	58	39	Nextera DNA Exome	881,277	79	25	92	-	-	-	-
CANIS0259	Illumina-DNA-Prep	4,328,561	6	40	16	Nextera DNA Exome	566,683	57	19	85	-	-	-	-
CANIS0278	Illumina-DNA-Prep	8,497,448	13	67	35	Nextera DNA Exome	572,487	59	20	87	-	-	-	-
CANIS0295	Illumina-DNA-Prep	9,391,412	13	71	37	Nextera DNA Exome	790,743	79	27	93	SQK LSK109 library prep	209,400	12	36
CANIS0298	Illumina-DNA-Prep	10,903378	17	69	40	Nextera DNA Exome	819,128	77	22	87	-	-	-	-
CANIS0304	Nextera-DNA-Prep	8,786369	15	68	40	Nextera DNA Exome	594,938	58	23	90	-	-	-	-
CANIS0342	Nextera-DNA-Prep	10,767,293	17	72	40	Nextera DNA Exome	505,919	50	25	87	-	-	-	-
CANIS0360	Illumina-DNA-Prep	5,600,141	9	52	29	Nextera DNA Exome	708,873	76	29	92	-	-	-	-
CANIS0363	Illumina-DNA-Prep	8,222,187	14	63	37	Nextera DNA Exome	792,584	83	27	93	SQK LSK109 library prep	196,481	12	31
CANIS0369	Nextera-DNA-Prep	13,720,032	22	86	40	Nextera DNA Exome	691,589	64	24	82	-	-	-	-
CANIS0376	Illumina-DNA-Prep	8,703,404	12	74	37	Nextera DNA Exome	177,934	30	24	65	-	-	-	-
CANIS0419	Nextera-DNA-Prep	9,928,408	17	98	41	Nextera DNA Exome	531,336	55	20	87	-	-	-	-
CANIS0461	Nextera-DNA-Prep	8,322,400	14	67	39	Nextera DNA Exome	340,853	37	19	80	-	-	-	-
CANIS0501	Illumina-DNA-Prep	7,319,097	13	64	39	Nextera DNA Exome	694,615	75	30	92	-	-	-	-
CANIS0540	Illumina-DNA-Prep	7,959,917	14	66	38	Nextera DNA Exome	656,907	63	27	91	-	-	-	-
CANIS0567	Nextera-DNA-Prep	9,524,504	16	58	39	Nextera DNA Exome	543,182	45	19	85	-	-	-	-
CANIS0591	Illumina-DNA-Prep	8,819,247	12	41	35	Nextera DNA Exome	606,045	62	22	85	-	-	-	-
CANIS0630	Nextera-DNA-Prep	9,438,745	16	90	38	Illumina DNA Prep with Enrichment	563,814	83	51	84	-	-	-	-
CANIS0664	Illumina-DNA-Prep	11,057,346	15	78	37	Nextera DNA Exome	880,939	81	32	94	SQK LSK109 library prep	127,364	8	20
CANIS0686	Illumina-DNA-Prep	8,607,509	12	40	34	Nextera DNA Exome	571,960	58	39	75	-	-	-	-
CANIS0688	Illumina-DNA-Prep	9,795,141	12	53	24	Nextera DNA Exome	292,213	39	29	67	-	-	-	-
CANIS0700	Nextera-DNA-Prep	7,556,693	14	77	40	Nextera DNA Exome	519,874	48	21	85	-	-	-	-
CANIS0705	Illumina-DNA-Prep	7,241,520	11	68	34	Nextera DNA Exome	819,291	91	35	94	SQK LSK109 library prep	151,730	11	32
CANIS0708	Illumina-DNA-Prep	5,906,851	10	54	32	Nextera DNA Exome	597,745	58	18	84	-	-	-	-
CANIS0781	Illumina-DNA-Prep	7,952,665	9	64	30	Nextera DNA Exome	278,520	44	30	72	SQK LSK109 library prep	157,784	11	35
CANIS0864	Nextera-DNA-Prep	9,430,010	16	80	40	Nextera DNA Exome	619,172	60	26	91	-	-	-	-
GOU044	Nextera-DNA-Prep	8,287,525	14	65	38	Illumina DNA Prep with Enrichment	566,886	85	43	77	-	-	-	-
MD-22	Nextera-DNA-Prep	6,940,271	12	59	38	Illumina DNA Prep with Enrichment	348,789	66	35	84	-	-	-	-
NA24385	Illumina-DNA-Prep	8,696,533	15	85	40	Illumina DNA Prep with Enrichment	828,536	94	109	90	-	-	-	-
SOK9	Nextera-DNA-Prep	4,160,960	7	35	10	Illumina DNA Prep with Enrichment	439,254	72	45	83	-	-	-	-
WAZ118	Nextera-DNA-Prep	7,666,855	13	59	37	Illumina DNA Prep with Enrichment	489,425	92	48	88	-	-	-	-

Table S2. Per sample Y chromosome haplogroup classifications for the five tools based on short-read data. The retrieved classification that matches the consensus is shown in bold. A) WGS and B) WES datasets.

Samples	Consensus haplogroup	Concordance rate (WGS)	Consensus haplogroup corrected	Whole-Genome Sequencing						Whole-Exome Sequencing					
				YLeaf (BAM)	Haplogrouper (BAM)	LineageTracker (BAM)	LineageTracker (VCF)	clean_tree_v2 (BAM)	pathPhynder (BAM)	YLeaf (BAM)	Haplogrouper (BAM)	LineageTracker (BAM)	LineageTracker (VCF)	clean_tree_v2 (BAM)	pathPhynder (BAM)
ATTA022	E1b1a1a1a2a1a3b1a1a19	100.00	E1b1a1a1a2a1a3b1a1a19	**E1b1a1a1a2a1a3b1a1a19**	**E1b1a1a1a2a1a3b1a1a19**	**E1b1a1a1a2a1a3b1a1a19**	**E1b1a1a1a2a1a3b1a1a19**	**E1b1a1a1a2a1a3b1a1a19**	**E1b1a1a1a2a1a3b1a1a19**	**E1b1a1a1a2a1a3b1a1a19**	**E1b1a1a1a2a1a3b1a1a19**	No Data	**E1b1a1a1a2a1a3b1a1a19**	**E1b1a1a1a2a1a3b1a1a19**	**E1b1a1a1a2a1a3b1a1a19**
BLA-050	R1b1b1b2a2a1	100.00	R1b1b1b2a2a1	**R1b1b1b2a2a1**	**R1b1b1b2a2a1**	**R1b1b1b2a2a1**	**R1b1b1b2a2a1**	**R1b1b1b2a2a1**	**R1b1b1b2a2a1**	**R1b1b1b2a2a1**	R1	No Data	P1xK2b2a	**R1b1b1b2a2a1**	**R1b1b1b2a2a1**
CAM-GFF-117	E1b1a1a1a1c2	33.33	E1b1a1a1a1c2	**E1b1a1a1a1c2**	**E1b1a1a1a1c2a**	**E1b1a1a1a1c2c3a1b**	**E1b1a1a1a1c2a**	**E1b1a1a1a1c2c3a1b**	I1a2a1a1a2b3a	**E1b1a1a1a1c2**	E1b1a1a1	No Data	E1b1a1a1	**E1b1a1a1a1c2**	I1a2a1a1a2b3a
CAM-PYKL-042	E1b1a1a1a1c1a1a3a1a1	100.00	E1b1a1a1a1c1a1a3a1	**E1b1a1a1a1c1a1a3a1a1**	**E1b1a1a1a1c1a1a3a1a1**	**E1b1a1a1a1c1a1a3a1a1**	**E1b1a1a1a1c1a1a3a1a1**	**E1b1a1a1a1c1a1a3a1a1**	**E1b1a1a1a1c1a1a3a1a1**	**E1b1a1a1a1c1a1a3a1**	E1b1a1a1a1c1a1	No Data	E1b1a1a1a1c1a1	**E1b1a1a1a1c1a1a3a1**	**E1b1a1a1a1c1a1a3a1**
CAM-PYKL-083	E1b1a1a1a1c1a1a3	100.00	E1b1a1a1a1c1a1a	**E1b1a1a1a1c1a1a3**	**E1b1a1a1a1c1a1a3**	**E1b1a1a1a1c1a1a3**	**E1b1a1a1a1c1a1a3**	**E1b1a1a1a1c1a1a3**	**E1b1a1a1a1c1a1a3**	**E1b1a1a1a1c1a1a**	E1b1a1a1a1c1a1	No Data	E1b1a1a1a1c1a1	**E1b1a1a1a1c1a1a**	**E1b1a1a1a1c1a1a**
CAM-PYKL-102	E1b1a1a1a2a1a3a2b1	50.00	E1b1a1a1a2a1a3a	**E1b1a1a1a2a1a3a2b**	**E1b1a1a1a2a1a3a2**	**E1b1a1a1a2a1a3a2b1**	**E1b1a1a1a2a1a3a2**	**E1b1a1a1a2a1a3a2b**	**E1b1a1a1a2a1a3a2b1**	**E1b1a1a1a2a1a3a**	E1b1a1a1	No Data	E1b1a1a1	**E1b1a1a1a2a1a3a**	**E1b1a1a1a2a1a3a**
CAM-PYKL-140	E1b1a1a1a1c1a1a3	100.00	E1b1a1a1a1c1a1a3	**E1b1a1a1a1c1a1a3**	**E1b1a1a1a1c1a1a3**	**E1b1a1a1a1c1a1a3**	**E1b1a1a1a1c1a1a3**	**E1b1a1a1a1c1a1a3**	**E1b1a1a1a1c1a1a3**	E1b1a1a1a1c1a1a3a1b1b1b1a2	E1b1a1a1	No Data	E1b1a1a1	E1b1a1a1a1c1a1a3a1b1b1b1a2	**E1b1a1a1a1c1a1a3**
CAM-PYKL-196	E1b1a1a1a2a1a3b1d1c1a	100.00	E1b1a1a1a2a1a3b1d1c1a	**E1b1a1a1a2a1a3b1d1c1a**	**E1b1a1a1a2a1a3b1d1c1a**	**E1b1a1a1a2a1a3b1d1c1a**	**E1b1a1a1a2a1a3b1d1c1a**	**E1b1a1a1a2a1a3b1d1c1a**	**E1b1a1a1a2a1a3b1d1c1a**	**E1b1a1a1a2a1a3b1d1c1a**	E1b1a1a1	No Data	E1b1a1a1	**E1b1a1a1a2a1a3b1d1c1a**	**E1b1a1a1a2a1a3b1d1c1a**
CANIS0005	J2a1	83.33	J2a1	**J2a1**	**J2a1**	**J2a1b3**	**J2a1**	**J2a1**	**J2a1**	**J2a1**	J2a	No Data	J2	**J2a1**	**J2a1**
CANIS0018	R1b1a1b1a1a2a5	83.33	R1b1a1b1a1a	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2c1a4b4a1a**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a**	R1	No Data	P1xK2b2a	**R1b1a1b1a1a**	**R1b1a1b1a1a**
CANIS0025	J1a2a1a2d2b2a1c	33.33	J1a2a1a2d2b2	**J1a2a1a2d2b2a1c**	J1a2a1a2b	**J1a2a1a2d2b2c1c1c4b1c3a**	**J1a2a1a2d2b2a1c**	J1a2a1a2d2b2a1c1	J1a2a1a2d2b2	J1a2a1a2d2b2c1c4i2a2a1c1	J	No Data	J	**J1a2a1a2d2b2c1c4i2a2a1c1**	**J1a2a1a2d2b2**
CANIS0031	R1b1a1b1a1a2a	66.67	R1b1a1b1a1a2a	**R1b1a1b1a1a2a**	**R1b1a1b1a1a2a**	R1b1a1b1a1a2c1a2a2d1a	**R1b1a1b1a1a2a**	R1b1a1b1a1a	**R1b1a1b1a1a2a**	R1b1a1b1a1a2c1a4a1	R1	No Data	P1xK2b2a	R1b1a1b1a1a2c1a4a1	R1b1a1b1a1a2c1a1a1a1a1
CANIS0057	R1b1a1b1a1a1c1a1	100.00	R1b1a1b1a1a1c1a1	**R1b1a1b1a1a1c1a1**	**R1b1a1b1a1a1c1a1**	**R1b1a1b1a1a1c1a1**	**R1b1a1b1a1a1c1a1**	**R1b1a1b1a1a1c1a1**	**R1b1a1b1a1a1c1a1**	R1b1a1b1a1a2b1c1a1a1	R1	No Data	P1xK2b2a	R1b1a1b1a1a2b1c1a1a1	R1b1a1b1a1a2b1c1a1a1
CANIS0097	J2a1a1a2d2b2b2b1	66.67	J2a1a1a2d2b2b	**J2a1a1a2b2a2b2b** 1	**J2a1a1a2b2a2b**	**J2a1a1a2b2a2b2b** 1	**J2a1a1a2b2a2b2b** 1	**J2a1a1a2b2a2b2b** 1	**J2a1a1a2b2a2b2b**	**J2a1a1a2b2a2b**	J2a1a	No Data	J2a1a	**J2a1a1a2b2a2b**	**J2a1a1a2b2a2b**
CANIS0123	R1a1a1b2a2a3	33.33	R1a1a1	**R1a1a1b2a2a3**	**R1a1a1b2a2**	**R1a1a1b2a1a2c1a**	**R1a1a1b2a2a3b**	**R1a1a1b2a2a3**	**R1a1a1b2a2a3b**	**R1a1a1**	R1	No Data	I1a2a1a1a2a2	**R1a1a1**	**R1a1a1**
CANIS0140	G2a2b2a1a1c1a1a2a1a1	100.00	G2a2b2a1a1c1a	**G2a2b2a1a1c1a1a2a1a1**	**G2a2b2a1a1c1a1a2a1a1**	**G2a2b2a1a1c1a1a2a1a1**	**G2a2b2a1a1c1a1a2a1a1**	**G2a2b2a1a1c1a1a2a1a1**	**G2a2b2a1a1c1a1a2a1a1**	**G2a2b2a1a1c1a1**	G2a2b2a1a1c1a1	No Data	G2a2b2a	**G2a2b2a1a1c1a1**	**G2a2b2a1a1c1a1**
CANIS0141	J2b2a1a1a1a1a1a1c	83.33	J2b2a1a1a1a1a1c	**J2b2a1a1a1a1a1c**	J2b2a1a1a1a1a	**J2b2a1a1a1a1a1a1c**	**J2b2a1a1a1a1a1c**	**J2b2a1a1a1a1a1c**	**J2b2a1a1a1a1a1c**	**J2b2a1a1a1a1a1c**	J2	No Data	J2	**J2b2a1a1a1a1a1a1c**	J2b2a1a1a1a1a1a1
CANIS0166	R1a1a1b1a2b3a3a1a1	83.33	R1a1a1b1a2b3a3a1	**R1a1a1b1a2b3a3a1a1**	**R1a1a1b1a2b3a3a1a1**	**R1a1a1b1a2b3a3a1a1**	**R1a1a1b1a2b3a3a1a1**	**R1a1a1b1a2b3a3a1a1**	**R1a1a1b1a2b3a3a1a1**	R1a1a1b1a2b3a3a1a	R1	No Data	I1a2a1a1a2a2	**R1a1a1b1a2b3a3a1a1**	R1a1a1b1a2b3a3a1a
CANIS0170	I1a3a1a1	66.67	I1a3a1a1	**I1a3a1a1**	**I1a3a1a1**	**I1a3a1a1b**	**I1a3a1a1b**	**I1a3a1a1b**	**I1a3a1a1**	**I1a3a1a1**	I1	No Data	I	**I1a3a1a1**	I1
CANIS0248	T1a2a	100.00	T1a2a	**T1a2a**	**T1a2a**	**T1a2a**	**T1a2a**	**T1a2a**	**T1a2a**	**T1a2a**	T1a	No Data	TorK1b	**T1a2a**	**T1a2a**
CANIS0259	E1b1b1a1a1c1a	50.00	E1b1b1a1a1c	**E1b1b1a1a1c1a**	**E1b1b1a1a1c1a**	E1b1b1a1a1c1b1a1a1	E1b1b1a1a1c1a	E1b1b1a1a1c1b1d	E1b1b1a1a1c1a	**E1b1b1a1a1c1a**	E1b1b1	No Data	E1b1b	E1b1b1a1a1c1a	E1b1b1a1a1c1b
CANIS0278	J1a2a1a2d2b2b2c4i2a2a5a1e2b1	66.67	J1a2a1a2d2b2b2c4i2a2a5a1e2b1	**J1a2a1a2d2b2b2c4e52a2a5a1e2b1**	J1a2a1a2b	**J1a2a1a2d2b2b2c4e52a2a5a1e2b1**	**J1a2a1a2d2b2b2c4e52a2a5a1e2b1**	**J1a2a1a2d2b2b2c4e52a2a5a1e2b1**	**J1a2a1a2d2b2b2c4e52a2a5a1e2b1**	**J1a2a1a2d2b2b2c4e52a2a5a**	J	No Data	J	**J1a2a1a2d2b2b2c4e52a2a5a**	J1a2a1a2d2b2b2c4i2a2a
CANIS0295	R1b1a1b1a1a2a1a1a1	83.33	R1b1a1b1a1a2a1a1a	**R1b1a1b1a1a2a1a1a1**	**R1b1a1b1a1a2a1a1a1**	R1b1a1b1a1a2c1a2a2d1a	**R1b1a1b1a1a2a1a1a1**	**R1b1a1b1a1a2a1a1a1**	**R1b1a1b1a1a2a1a1a1**	**R1b1a1b1a1a2a1a1a1**	R1	No Data	P1xK2b2a	**R1b1a1b1a1a2a1a1a**	**R1b1a1b1a1a2a1a1a**
CANIS0298	G2a2b1b1a1b1b1e	50.00	G2a2b1b1a1b1e	**G2a2b1b1a1b1e**	G2a2b1b1a1b1	G2a2b1b1a1b1a1b	**G2a2b1b1a1b1e**	**G2a2b1b1a1b1e**	**G2a2b1b1a1b1e**	G2a2b1b1a3	R1	No Data	.	**G2a2b1b1a1b1e**	**G2a2b1b1a1b1e**
CANIS0304	R1b1a1b1a1a2c1a1d3a	83.33	R1b1a1b1a1a2c1a1d3	**R1b1a1b1a1a2c1a1d3a**	**R1b1a1b1a1a2c1a1d3a**	R1b1a1b1a1a2c1a2a2d1a	**R1b1a1b1a1a2c1a1d3a**	**R1b1a1b1a1a2c1a1d3a**	**R1b1a1b1a1a2c1a1d3a**	**R1b1a1b1a1a2c1a1d3a**	I2a1a	No Data	P1xK2b2a	**R1b1a1b1a1a2c1a1d3a**	**R1b1a1b1a1a2c1a1d3a**
CANIS0342	I2a1a1a1a1a	50.00	I2a1a1a1	**I2a1a1a1a1a**	**I2a1a1a1a1a**	**I2a1a1a1a1a4c3a1**	**I2a1a1a1a1a1a1c5**	**I2a1a1a1a1a**	**I2a1a1a1a1a5**	**I2a1a1a1a1**	R1	No Data	I2a1a1a1	**I2a1a1a1a1**	**I2a1a1a1a1**
CANIS0360	R1b1a1b1a1a2a	33.33	R1b1a1b1a	**R1b1a1b1a1a2a**	**R1b1a1b1a1a2a**	**R1b1a1b1a1a2a2d1a**	**R1b1a1b1a1a2a7**	**R1b1a1b1a1a**	**R1b1a1b1a1a2a7**	**R1b1a1b1a1a**	G2a2b2a	No Data	P1xK2b2a	**R1b1a1b1a1a**	**R1b1a1b1a1a2a7**
CANIS0363	G2a2b2a1a1f	66.67	G2a2b2a1f	**G2a2b2a1a1f**	**G2a2b2a1a1f**	G2a2b2a1a1c2b1b4	**G2a2b2a1a1f**	**G2a2b2a1a1f**	**G2a2b2a1a1f**	G2a2b2a1	-	No Data	G2a2b2a	**G2a2b2a1a1f**	**G2a2b2a1a1f**
CANIS0369	E1b1b1b1a1c	83.33	E1b1b1b1a1	**E1b1b1b1a1c**	**E1b1b1b1a1c**	**E1b1b1b1a1c**	**E1b1b1b1a1c**	**E1b1b1b1a1c**	**E1b1b1b1a1c**	**E1b1b1b1a1**	**E1b1b1b1a1**	No Data	E1b1b	**E1b1b1b1a1**	**E1b1b1b1a1**
CANIS0376	R1b1a1b1a1a2a5	83.33	R1b1a1b1a1a	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2c1a2a2d1a**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a**	R1	No Data	P1xK2b2a	**R1b1a1b1a1a**	**R1b1a1b1a1a**
CANIS0419	R1b1a1b1a1a2a1a1a1a	83.33	R1b1a1b1a1a2a1a1a	**R1b1a1b1a1a2a1a1a1a**	**R1b1a1b1a1a2a1a1a1a**	R1b1a1b1a1a2c1a2a2d1a	**R1b1a1b1a1a2a1a1a1a**	**R1b1a1b1a1a2a1a1a1a**	**R1b1a1b1a1a2a1a1a1a**	R1b1a1b1a1a2b1c2b1a	R1	No Data	P1xK2b2a	R1b1a1b1a1a2b1c2b1a	**R1b1a1b1a1a2a1a1a**
CANIS0461	R1b1a1b1a1a2c1a1i1	83.33	R1b1a1b1a1a2c1a1i	**R1b1a1b1a1a2c1a1i1**	**R1b1a1b1a1a2c1a1i1**	R1b1a1b1a1a2c1a2a2b1b1a3	**R1b1a1b1a1a2c1a1i1**	**R1b1a1b1a1a2c1a1i1**	**R1b1a1b1a1a2c1a1i1**	R1b1a1b1a1a2c1a1a1a1a1a1a1a1a1a2	R1	No Data	P1xK2b2a	R1b1a1b1a1a2c1a1a1a1a1a1a1a1a1a2	**R1b1a1b1a1a2c1a1i1**
CANIS0501	R1b1a1b1a1a1c1a1	100.00	R1b1a1b1a1a	**R1b1a1b1a1a1c1a1**	**R1b1a1b1a1a1c1a1**	**R1b1a1b1a1a1c1a1**	**R1b1a1b1a1a1c1a1**	**R1b1a1b1a1a1c1a1**	**R1b1a1b1a1a1c1a1**	**R1b1a1b1a1a**	R1	No Data	P1xK2b2a	**R1b1a1b1a1a**	**R1b1a1b1a1a**
CANIS0540	R1b1a1b1a1a2c1a1a1e	66.67	R1b1a1b1a1a2c1a1a	**R1b1a1b1a1a2c1a1a**	**R1b1a1b1a1a2c1a1a**	R1b1a1b1a1a2c1a4i4e1a	**R1b1a1b1a1a2c1a1a**	R1b1a1b1a1a2c	**R1b1a1b1a1a2c1a1a**	**R1b1a1b1a1a2c1a1a**	R1	No Data	P1xK2b2a	R1b1a1b1a1a1c2i2b2b1a1a1b1	R1b1a1b1a1a
CANIS0567	E1b1b1b1a1	83.33	E1b1b1b1a1	**E1b1b1b1a1**	**E1b1b1b1a1**	**E1b1b1b1a1b3a**	**E1b1b1b1a1**	**E1b1b1b1a1**	**E1b1b1b1a1**	**E1b1b1b1a1**	**E1b1b1b1a1**	No Data	E1b1b	E1b1b1b1a2	**E1b1b1b1a1**
CANIS0591	R1b1a1b1a1a2c1a3	83.33	R1b1a1b1a1a2c	**R1b1a1b1a1a2c1a3**	**R1b1a1b1a1a2c1a3**	**R1b1a1b1a1a2c1a5d**	**R1b1a1b1a1a2c1a3**	**R1b1a1b1a1a2c1a3**	**R1b1a1b1a1a2c1a3**	**R1b1a1b1a1a2c1a4b8**	R1	No Data	P1xK2b2a	**R1b1a1b1a1a2c1a4b8**	**R1b1a1b1a1a2c**
CANIS0620	R1a1a1b1a2b3a3a1a1	83.33	R1a1a1b1a2b3a3a1a	**R1a1a1b1a2b3a3a1a1**	**R1a1a1b1a2b3a3a1a1**	**R1a1a1b1a2b3a3a1a1**	**R1a1a1b1a2b3a3a1a1**	**R1a1a1b1a2b3a3a1a1**	**R1a1a1b1a2b3a3a1a1**	**R1a1a1b1a2b3a3a1a**	R1	No Data	I1a2a1a1a2a2	**R1a1a1b1a2b3a3a1a**	**R1a1a1b1a2b3a3a1a**
CANIS0664	E1b1b1b1a1c	100.00	E1b1b1b1a1	**E1b1b1b1a1c**	**E1b1b1b1a1c**	**E1b1b1b1a1c**	**E1b1b1b1a1c**	**E1b1b1b1a1c**	**E1b1b1b1a1c**	**E1b1b1b1a1**	**E1b1b1b1a1**	No Data	E1b1b	**E1b1b1b1a1**	**E1b1b1b1a1**
CANIS0686	R1b1a1b1a1a2a5	83.33	R1b1a1b1a1a2	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2c1a2a2d1a**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2**	R1	No Data	P1xK2b2a	**R1b1a1b1a1a2**	**R1b1a1b1a1a2**
CANIS0688	R1b1a1b1a1a2a	66.67	R1b1a1b1a	**R1b1a1b1a1a2a**	**R1b1a1b1a1a2a**	**R1b1a1b1a1a2c1a1f1b1b**	**R1b1a1b1a1a2a**	**R1b1a1b1a1a**	**R1b1a1b1a1a2a**	**R1b1a1b1a1a**	R1	No Data	P1xK2b2a	**R1b1a1b1a1a2c1a2a2a1b1**	**R1b1a1b1a1a**
CANIS0700	R1b1a1b1a1a2a5	83.33	R1b1a1b1a1a2	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2c1a2a2d1a**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2a5**	**R1b1a1b1a1a2**		No Data	P1xK2b2a	**R1b1a1b1a1a2c1a2a2a1d1a**	**R1b1a1b1a1a2**
CANIS0705	R1b1a1b1a1a2b1a1	33.33	R1b1a1b1a1a2	**R1b1a1b1a1a2b1a1**	**R1b1a1b1a1a2b**	**R1b1a1b1a1a2c1a2a2d1a**	**R1b1a1b1a1a2b**	**R1b1a1b1a1a2**	**R1b1a1b1a1a2**	**R1b1a1b1a1a2a1a1**	R1	No Data	P1xK2b2a	R1b1a1b1a1a2c2b1a1a1	**R1b1a1b1a1a2**
CANIS0708	R1b1a1b1a1a2a1b1a1	50.00	R1b1a1b1a1a2a1b1a1	**R1b1a1b1a1a2a1b1a1**	**R1b1a1b1a1a2a1b1a1**	R1b1a1b1a1a2c1a2a2d1a	**R1b1a1b1a1a2a1b1a1**	**R1b1a1b1a1a2a1b1a1**	**R1b1a1b1a1a2a1b1a1a1a**	**R1b1a1b1a1a2a1b1a1**	R1	No Data	P1xK2b2a	**R1b1a1b1a1a2a1b1a1**	E1b1a1a1a2a1a3b1d1c1b
CANIS0781	T1a1a1b2b2b1a1a	100.00	T1a1a1b2b2b	**T1a1a1b2b2b1a1a**	**T1a1a1b2b2b1a1a**	**T1a1a1b2b2b1a1a**	**T1a1a1b2b2b1a1a**	**T1a1a1b2b2b1a1a**	**T1a1a1b2b2b1a1a**	**T1a1a1b2b2b**	**T1a1a1b2b2b**	No Data	TorK1b	**T1a1a1b2b2b**	**T1a1a1b2b2b**
CANIS0864	R1b1a1b1a1a2a1a1a1	83.33	R1b1a1b1a1a2a1a1	**R1b1a1b1a1a2a1a1a1**	**R1b1a1b1a1a2a1a1a1**	R1b1a1b1a1a2c1a2a2d1a	**R1b1a1b1a1a2a1a1a1**	**R1b1a1b1a1a2a1a1a1**	**R1b1a1b1a1a2a1a1a1**	**R1b1a1b1a1a2a1a1a1**	R1	No Data	P1xK2b2a	**R1b1a1b1a1a2a1a1a1**	**R1b1a1b1a1a2a1a1a1**
GOU004	E1b1a1a1a2a1a3b1a	100.00	E1b1a1a1a2a1a3b1a	**E1b1a1a1a2a1a3b1a**	**E1b1a1a1a2a1a3b1a**	**E1b1a1a1a2a1a3b1a**	**E1b1a1a1a2a1a3b1a**	**E1b1a1a1a2a1a3b1a**	**E1b1a1a1a2a1a3b1a**	**E1b1a1a1a2a1a3b1a**	E1b1a1a1	No Data	E1b1a1a1	**E1b1a1a1a2a1a3b1a**	**E1b1a1a1a2a1a3b1a**
MD-22	E1a1a2b	100.00	E1a1	**E1a1a2b**	**E1a1a2b**	**E1a1a2b**	**E1a1a2b**	**E1a1a2b**	**E1a1a2b**	**E1a1**	E	No Data	E	**E1a1**	**E1a1**
NA3435	J1a2a1a2c1a1a1	50.00	J1a2a1a2c1a1a1	J1a2a1a2c1a1	**J1a2a1a2c1a1a1**	**J1a2a1a2c1a1a1**	**J1a2a1a2c1a1a1**	J1a2a1a2c1a1a	**J1a2a1a2c1a1a1**	**J1a2a1a2c1a1a1**	J1a2a1a2c1a1	No Data	J1a2a1a2c1a1	**J1a2a1a2c1a1a1**	**J1a2a1a2c1a1a1**
SGK9	R1b1b2a2b	66.67	R1b1b2a2b	R1b1b2a2	**R1b1b2a2b**	**R1b1b2a2b**	**R1b1b2a2b**	R1b1b2a2	**R1b1b2a2b**	R1b1b2a2	R1	No Data	P1xK2b2a	R1b1b2a2	**R1b1b2a2b**
WAZ110	E1b1a1a1c1b2a1	50.00	E1b1a1a1c1b2a	**E1b1a1a1c1b2a1**	**E1b1a1a1c1b2a1**	**E1b1a1a1c1b2a1b**	**E1b1a1a1c1b2a1b**	**E1b1a1a1c1b2a1**	**E1b1a1a1c1b2a1b**	**E1b1a1a1c1b2a**	E1b1a1a1	No Data	E1b1a1a1	**E1b1a1a1c1b2a**	**E1b1a1a1c1b2a**