Supplementary Information

De novo whole-genome assembly and resequencing resources for the roan (*Hippotragus equinus*), an iconic African antelope

Margarida Gonçalves,*[,†,1] Hans R. Siegismund,[‡] Bettine Jansen van Vuuren,[§] Klaus-Peter Koepfli,**[,††] Nuno Ferrand*[,†,§] & Raquel Godinho*[,†,§,1]

*CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus de Vairão, Portugal

†Departamento de Biologia, Faculdade de Ciências, Universidade do Porto

‡Department of Biology, University of Copenhagen, Denmark

§Centre for Ecological Genomics and Wildlife Conservation, Department of Zoology, University of Johannesburg, South Africa

**Smithsonian-Mason School of Conservation, Front Royal, VA 22630, USA

††Smithsonian Conservation Biology Institute, Center for Species Survival, National Zoological Park, Washington, DC 20008, USA

[1]Corresponding author: CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus de Vairão, Portugal. E-mail: mgoncalves@cibio.up.pt; rgodinho@cibio.up.pt

TABLES

TABLE S1. Genome assembly statistics calculated by Supernova v.2.1.1, using randomly selected reads to achieve 56x depth of coverage.

	Assembly run
Read number (M)	1212.03
Mean read length (bp)	139.50
Raw coverage (x)	60.70
Effective mean coverage (x)	39.70
Read two Q30 (%)	80.70
Median insert (bp)	335.00
Proper pairs (%)	90.32
Molecule length (Kbp)	21.57
Heterozygosity distance (bp)	636.00
Unbarcoded reads (%)	7.05
N50 reads per barcode (K)	0.94
Duplicates (%)	26.01
Phased (%)	48.13
Scaffolds >= 10Kbp (K)	2.20
N50 edge size (Kbp)	30.36
N50 contig size (Kbp)	239.59
N50 phased block size (Mbp)	1.40
N50 scaffold size (Mbp)	8.42
Base missing from scaffolds >= 10Kb (%)	4.10
Assembly size (scaffolds >= 10Kb) (Gbp)	2.56
Estimated genome size (Gbp)	3.01

x – covered reads; K – thousand; M – million; bp – base-pairs; Kbp – thousand base-pairs; Mbp – million base-pairs; Gbp – billion base-pairs.
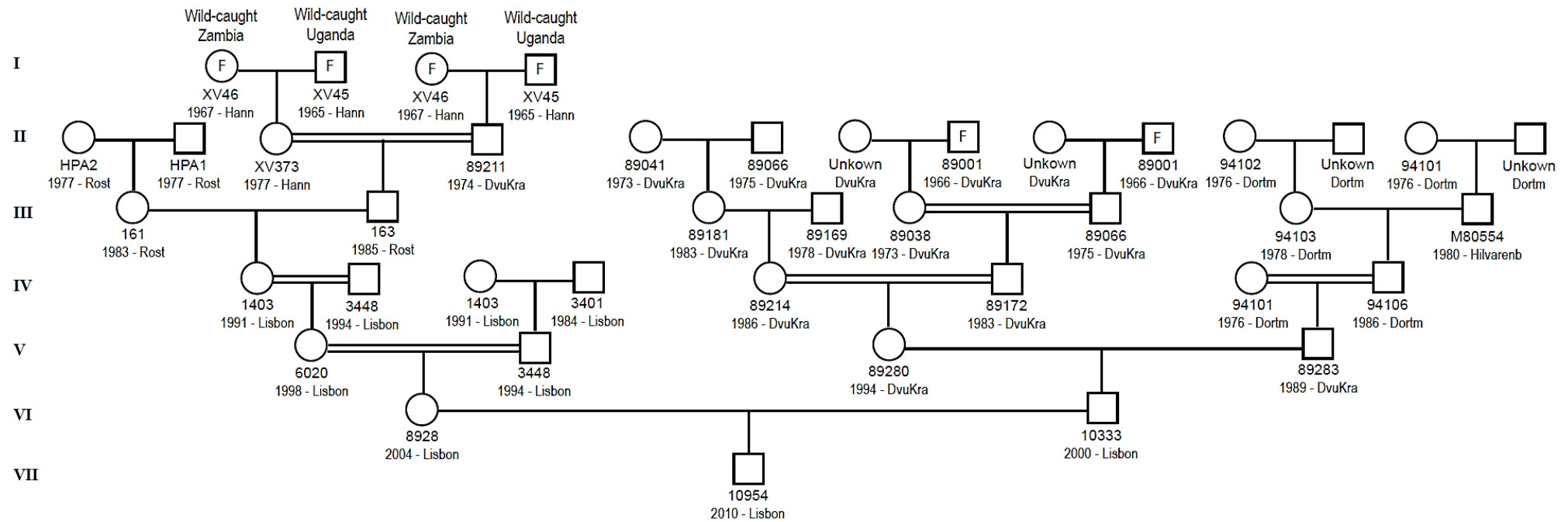
TABLE S2. Genome annotation quality assessment using GenomeQC tool, comparing roan antelope with cow reference genome (BosTau9), sable antelope genome and the scimitar-horned oryx genome.

	Roan antelope	Domestic cow	Sable antelope	Scimitar-horned oryx
Number of gene models	30 622	30 589	23 846	28 559
Minimum gene length	56	45	6	6
Maximum gene length	1 118 330	2 433 917	163 506	162 991
Average gene length	37 281.0	39 806.5	16 709.8	16 128.4
Number of transcripts	25 139	66 643	23 846	28 559
Average number of transcripts per gene model	2.0	2.2	1.0	1.0

TABLE S3. Alignment statistics of the roan antelope genome against the domestic cow reference genome (BosTau9), per chromosome. A complete description of roan scaffolds alignment to each domestic cow chromosome is provided in Table S4.

	Identity (%)	Mapped bp (N)	Mapped scaffolds (N)	Mapping (%)
chr1	92.89	155 462 201	32	98.06
chr2	93.03	134 758 668	41	98.92
chr3	93.00	116 791 714	64	96.52
chr4	92.91	116 204 881	28	96.84
chr5	92.88	114 968 136	52	95.74
chr6	92.81	115 955 195	24	98.43
chr7	92.92	106 359 928	68	96.09
chr8	92.97	110 519 018	43	97,53
chr9	92.82	102 494 769	16	97.19
chr10	92.99	98 360 627	29	95.21
chr11	92.83	105 507 747	26	98.62
chr12	92.51	83 334 402	23	95.55
chr13	92.84	79 717 752	32	95.50
chr14	92.85	80 159 661	21	97.28
chr15	92.77	70 396 358	44	82.81
chr16	92.72	77 921 579	39	96.18
chr17	92.55	71 298 713	28	97.45
chr18	92.65	61 678 633	58	93.71
chr19	92.49	60 956 150	40	96.07
chr20	92.61	69 947 538	8	97.18
chr21	92.65	67 359 392	61	96.42
chr22	92.50	59 787 467	21	98.38
chr23	92.45	46 624 102	38	88.81
chr24	92.62	61 424 422	16	98.57
chr25	92.14	41 507 042	19	98.01
chr26	92.54	49 906 710	27	95.99
chr27	92.37	43 147 230	12	94.60
chr28	92.55	44 358 876	29	96.56
chr29	92.25	48 728 272	27	95.36
chrX	93.30	115 206 517	261	82.88

FIGURES

I

Wild-caught Zambia — F XV46 1967 - Hann
Wild-caught Uganda — F XV45 1965 - Hann
Wild-caught Zambia — F XV46 1967 - Hann
Wild-caught Uganda — F XV45 1965 - Hann

II

HPA2 1977 - Rost
HPA1 1977 - Rost
XV373 1977 - Hann
89211 1974 - DvuKra
89041 1973 - DvuKra
89066 1975 - DvuKra
Unkown DvuKra
F 89001 1966 - DvuKra
Unkown DvuKra
F 89001 1966 - DvuKra
94102 1976 - Dortm
Unkown Dortm
94101 1976 - Dortm
Unkown Dortm

III

161 1983 - Rost
163 1985 - Rost
89181 1983 - DvuKra
89169 1978 - DvuKra
89038 1973 - DvuKra
89066 1975 - DvuKra
94103 1978 - Dortm
M80554 1980 - Hilvarenb

IV

1403 1991 - Lisbon
3448 1994 - Lisbon
1403 1991 - Lisbon
3401 1984 - Lisbon
89214 1986 - DvuKra
89172 1983 - DvuKra
94101 1976 - Dortm
94106 1986 - Dortm

V

6020 1998 - Lisbon
3448 1994 - Lisbon
89280 1994 - DvuKra
89283 1989 - DvuKra

VI

8928 2004 - Lisbon
10333 2000 - Lisbon

VII

10954 2010 - Lisbon

Lisbon - Lisbon Zoo (Portugal)
DvuKra - Safari Park Dvur Králové (Czech Republic)
Dortm - Dortmund Zoo (Germany)
Rost - Rostock Zoo (Germany)
Hilvarenb - Hilvarenbreek Zoo (Netherlands)
Hann - Hannover Zoo (Germany)

F - Zoo Founder

FIGURE S1. Pedigree of roan antelope individual ID:10954, used to produce the assembled genome. Six generations are depicted, indicated by Roman numbers, with generation I made up of wild-caught individuals to become zoo founders (F). Males are indicated by squares and females by circles. Consanguineous mattings are indicated by double line connections between male and female. Each individual is identified by an ID code, year of birth and zoo name. Adapted from pedigree made available by the Lisbon Zoo.
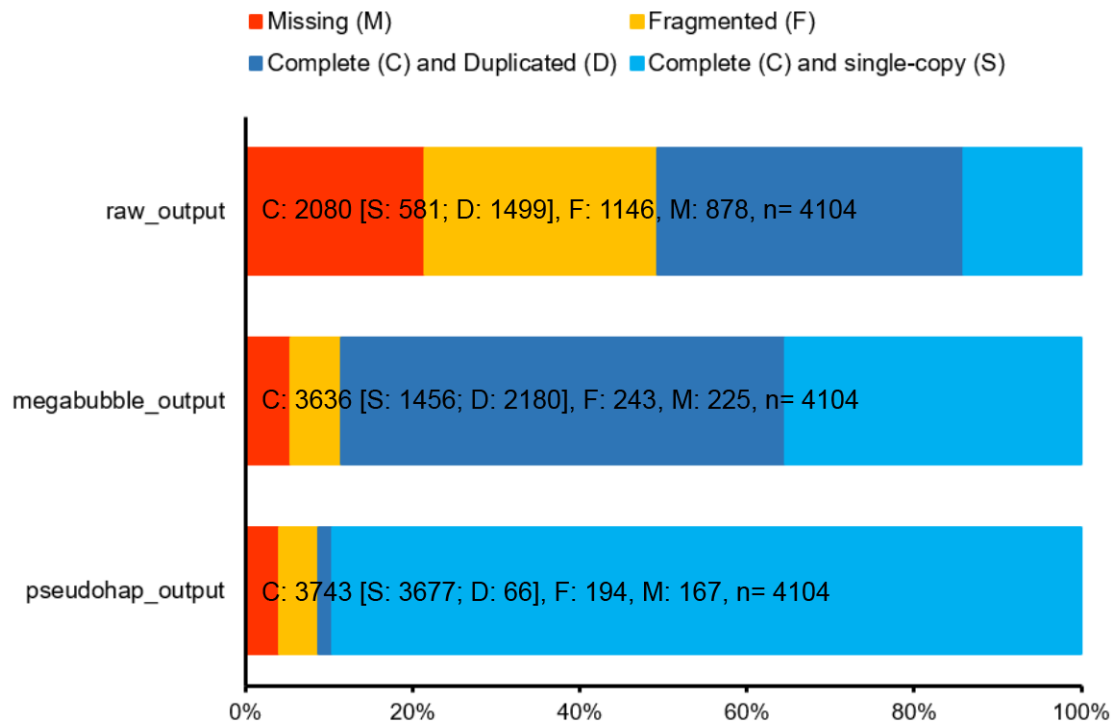
FIGURE S2. Plot comparing gene completeness of the roan genome assembly according to BUSCO v.3.0.2 percentage scores, among the three Supernova assembly output styles: raw; megabubble; and pseudohaplotype. BUSCO scores were estimated using the Mammalia OrthoDB v9 gene set, containing 4,104 genes. Total number of genes falling into the four categories, according to the legend and colour code.