**Supporting Information**

**Supporting Figures**

**Figure S1** Distribution of the *k*-mer frequency for the roe deer genomic DNA sequence data set.

**Supporting Tables**

**Table S1** Detailed information of the *de novo* deer assembly.

|  |  |  |
| --- | --- | --- |
| **Assembly** | **Deer\_SOAPdenovo\_k43.scafSeq** | **Deer\_SOAPdenovo\_k63.scafSeq** |
| Number of scaffolds | 1406571 | 3088511 |
| Total size of scaffolds | 2426270293 | 2785377831 |
| Total scaffold length as percentage of assumed genome size | 69 | 79 |
| Longest scaffold | 131768 | 151776 |
| Shortest scaffold | 100 | 100 |
| Number of scaffolds > 1K nt | 317505 | 314210 |
| Percentage of scaffolds > 1K nt | 23 | 10 |
| Number of scaffolds > 10K nt | 65469 | 71540 |
| Percentage of scaffolds > 10K nt | 5 | 2 |
| Number of scaffolds > 100K nt | 21 | 49 |
| Mean scaffold size | 1725 | 902 |
| Median scaffold size | 202 | 127 |
| N50 scaffold length | 9822 | 10458 |
| N90 scaffold length | 870 | 183 |
| scaffold %A | 28 | 29 |
| scaffold %C | 20 | 21 |
| scaffold %G | 20 | 21 |
| scaffold %T | 28 | 29 |
| scaffold %N | 3 | 2 |

**Table S2** Primer sequences used in this study.

|  |  |  |  |
| --- | --- | --- | --- |
| **Genes** | **Primer designations** | **Primer sequences (5’→3’)** | **PCR product sizes [bp]** |
| **Fragment analyses** |
| *AMELX/Y* | Ex5\_F | ACACCACCAGCCAAACCTCC | X: 287 |
|  | Ex5\_R | CCCGCTTGGTCTTGTCTGTTGC 1 | Y: 224 |
| *SRY* | deerRG4\_F | ggtcaagcgacccatgaa(C/T)gcNtt 2 | 216 |
|  | deerRG7\_R | ggtcgatacttatagttcgggta(C/T)tt 2 |  |
|  | BOVIRBP\_F | tgtatgatcaccttctatgcttc 2 | 180 |
|  | BOVIRBP\_R | gctttaggtaatcatcagatagc 2 |  |
| **Sequence analyses\*** |
| *AR* | Ex1-1\_F | AGCCTGTTGAACTCTTCTCAGCA | 785 |
|  | Ex1-1\_R | TGCCTTACACAACTCCTTGGC |  |
|  | Ex1-2\_F | AGCAGCGGGAGAGCGAGGGAG | 490 |
|  | Ex1-2\_R | GGGGAAAGTTGTAGTATTCGCGA |  |
|  | Ex1-3\_F | ACTGCTGAGTATTCCCCTTTCAAGG | 713 |
|  | Ex1-3\_R | GGGGAGAGGACTGACTCTAGGCTT |  |
|  | Ex2\_F | GCAGTTTAATCCTGAAGATTCAGACTTC | 323 |
|  | Ex2\_R | CATAAAATCCCGAGTCCTGATAGAC |  |
|  | Ex3\_F | TTGGGCAGGATATGTCTCACAC | 323 |
|  | Ex3\_R | CTGGCCTCCTTAGAAGCATCA |  |
|  | Ex4\_F | CTCAAGCAGCTTTTCTAACAGGG | 451 |
|  | Ex4\_R | GGAAAGGAACAGATGTGGTGC |  |
|  | Ex5\_F | ATATAATCTGTTCTCCTTTTACTCCCCA | 354 |
|  | Ex5\_R | CATTGCCATGAACCAGGTCTGT |  |
|  | Ex6\_F | ACATCAAGCTCTTCTTGGAGAAACTG | 352 |
|  | Ex6\_R | AATATTCAGGAGCTGGCTTCTTCC |  |
|  | Ex7\_F | CAAGTTTGTGGTCAGAAAACTTGG | 299 |
|  | Ex7\_R | TGGCTCTATCAGGCTACTCTCCC |  |
|  | Ex8\_F | CCTGGAGACAATTTAATGCACATGT | 442 |
|  | Ex8\_R | ATTCCCCAAGGCACTGCAGAG |  |
| *DMRT1* | Ex1\_F | AGAACGTACGCGTCCTGCG | 469 |
|  | Ex1\_R | GGGCTTTACTCGAAAAATTCGG |  |
|  | Ex2\_F | GATTCTCAGACCTTCTCCCTGGA | 392 |
|  | Ex2\_R | GGCTCGGAACCATCCCA |  |
|  | Ex3\_F | AGGACGGAGGCCGTGTG | 422 |
|  | Ex3\_R | GCATGTGTGCGGTTTCAAATATC |  |
|  | Ex4\_F | TTATTCATGGGTGAGGACACTTCTAA | 320 |
|  | Ex4\_R | GTGCCAGCAGGTATCCCCT |  |
|  | Ex5\_F | TGTTGGAACCTAATTGCATACTGG | 315 |
|  | Ex5\_R | AAACCCTGTGTATAACGAAATAAGCC |  |
| *FGF9* | Ex1\_F | AGTTGGATTTACCTCGCCGAGT | 418 |
|  | Ex1\_R | AGCTGCGGTCCGAGCCTA |  |
|  | Ex2\_F | AAGATGGTTCCCTAGTCAGAGCTC | 273 |
|  | Ex2\_R | GAAAGAGTGATTCCCAGCAAATG |  |
|  | Ex3\_F | GGGTTAAAATCGGTATATTCTCACCA | 408 |
|  | Ex3\_R | GGACCCACCGCGTGAA |  |
| *FOXL2* | Ex1-1\_F | GTGAACGGACTTGTGCGCC | 523 |
|  | Ex1-1\_R | GGTAGTTGCCCTTCTCGAACATG |  |
|  | Ex1-2\_F | TCAGCCTCAACGAGTGCTTCAT | 515 |
|  | Ex1-2\_R | CAGACCGTTGTACGAGTTCACCA |  |
|  | Ex1-3\_F | CTCGTACGGGCCGTACTCG | 428 |
|  | Ex1-3\_R | AGAGCTAGGAGGCGCAGAG |  |
| *RSPO1* | Ex2\_F | TCTCCTGGGTAGTGCACCAGAT | 579 |
|  | Ex2\_R | ATCACAGAATGAAATGGCCGAA |  |
|  | Ex3\_F | CACATTCGTTCTGTAGGCTGAAAC | 392 |
|  | Ex3\_R | CTCTTGCAGCCCATCAACAGT |  |
|  | Ex4\_F | CGGCAGTGGCAGCTCCA | 351 |
|  | Ex4\_R | GGCTCTCTCCGCACGGAG |  |
|  | Ex5\_F | TATGCCCTCTGGACCATCAGG | 392 |
|  | Ex5\_R | GGTAGGAAACAGAGGCGATCAGA |  |
|  | Ex6\_F | GAGGAGGCCTGGAGGGTACC | 369 |
|  | Ex6\_R | ACAGCGGCACTGGACTCCTT |  |

**Table S2** Continued.

|  |  |  |  |
| --- | --- | --- | --- |
| **Genes** | **Primer designations** | **Primer sequences (5’→3’)** | **PCR product sizes [bp]** |
| **Sequence analyses\*** |
| *SOX3* | 5’UTR\_F | GTATATAAGAAGCCCGGCGAGC | 1470 |
|  | 3’UTR\_R | GCAACAGTCCCAGGCAAGC |  |
|  | Ex1-1\_F | CAGGCAAgAGtAGTGCGAACG | 512 |
|  | Ex1-1\_R | TTCACGTGTGTGTACGTGTCCA |  |
|  | Ex1-2\_F | CCAAGATGCACAACTCCGAGAT | 380 |
|  | Ex1-2\_R | GCTGCTCATGGTCGCGG |  |
|  | Ex1-3\_F | GTGAACGGCTGGGCCAA | 641 |
|  | 3’UTR\_R | GCAACAGTCCCAGGCAAGC |  |
| *SOX9* | Ex1\_F | CGCCTCGAGTGCCCGA | 574 |
|  | Ex1\_R | CGCCGCGGAGCCCA |  |
|  | Ex2\_F | TTCATCCGCAGGGTTTCCA | 618 |
|  | Ex2\_R | TGTCCCGATGCAGAGCCTC |  |
|  | Ex3-1\_F | GTTAATCATTCTGTGCCTTATCTCGG | 753 |
|  | EX3-1\_R | CTTGATGTGCGTGCGCTG |  |
|  | Ex3-2\_F | GTGTGGATGTCCAAGCAGCAG | 683 |
|  | Ex3-2\_R | ATGTCCGAAGGGTCGCGT |  |
| *SOX10* | Ex2-1\_F | CTTAGGGCTGTGCCTGAGTCG | 456 |
|  | Ex2-1\_R | CCGTTGACGCGCACCG |  |
|  | Ex2-2\_F | ATGACAAGTTCCCCGTGTGCA | 296 |
|  | Ex2-2\_R | ATTCCACCCGCGGGTAGAG |  |
|  | Ex3\_F | CTAGACCAGACAGCGCGACTCC | 471 |
|  | Ex3\_R | TCTCAAGTGCTGCTCCAGCCTT |  |
|  | Ex4\_F | CGTCTGAAAGTGAACCATCAGCCACT | 614 |
|  | Ex4\_R | GTAGTGGGGCAGACTGAGGGAGGTGTA |  |
| *SRY* | Ex1-1\_F | GGTAAGAACAACTTATTTAACAGCACGAT | 519 |
|  | Ex1-1\_R | TTGAAGAGTCTGCAGGAAGCAAT |  |
|  | Ex1-2\_F | GGTATGAGTGGAAAAGGCTTACAGAT | 489 |
|  | Ex1-2\_R | AGTCACAGGTGAAATTGTAGTAAAATTGAGA |  |
| *WT1* | Ex1\_F | GGCATCTGGGCCAAGTTAGG | 649 |
|  | Ex1\_R | ACCGGTTAAGAACTGTGGTCAGG |  |
|  | Ex2\_F | ACTGGACCCGGCTTCCCT | 335 |
|  | Ex2\_R | GCTGTGGGTTGGAAATTCCTG |  |
|  | Ex3\_F | TTGCGCCCTGCACAGC | 262 |
|  | Ex3\_R | CCGCCAGCTCGAGCATC |  |
|  | Ex4\_F | CATGTGGAGATCAGTGGGATGA | 360 |
|  | Ex4\_R | CACGCCTATTCTACAACTGTGCTCT |  |
|  | Ex5\_F | TCAGGGACCCAGTTCAGCA | 287 |
|  | Ex5\_R | GATTACCCGTCATCAGTCCTAACTC |  |
|  | Ex6\_F | CCCGGGAACTCTGAGAGGA | 307 |
|  | Ex6\_R | TGACACAGGAGCCAATGGTTAG |  |
|  | Ex7\_F | TGGTTTCAGGCATCGCATC | 355 |
|  | Ex7\_R | TGGCAGTGACAGCTAGAAGCAG |  |
|  | Ex8\_F | GTCCTTTCAGTGCCAGTAAGAGAAGT | 342 |
|  | Ex8\_R | TTTTAAAGAGAAGGTAAATCCCTAGCAA |  |
|  | Ex9\_F | AGGTATGGCTGGAAATGCCC | 285 |
|  | Ex9\_R | ATCTCTCACATCACAAGCAAATCC |  |
|  | Ex10\_F | AGGTGCTGCCATATCAAGGC | 391 |
|  | Ex10\_R | GACAGGCGAGTGAGGAGGAAT |  |
| **Real time PCR analyses** |
| *AR* | Ex2RT\_F | TTTGGAGACCACCAGGGACCA | 152 |
|  | Ex2RT\_R | CTTCAGCGGCTCTTTTAAAAAAGAC |  |
| *DMRT1* | Ex3RT\_F | GTCATCCAGGATATTCCTGCTGTC | 148 |
|  | Ex3RT\_R | TGGAGTACTGCGGGTAGTTGTAGAG |  |
|  | Ex4RT\_F | AACCCGGAGAACCTCCACG | 139 |
|  | Ex4RT\_R | CGCTCGTTTTGGCTTCCG |  |
| *FGF9* | Ex1RT\_F | GGGTCAGTCCGAAGCAGGAG | 142 |
|  | Ex1RT\_R | TCCCTGGATAGTACCGTTGGG |  |
|  | Ex2RT\_F | CAGGGATCCTGGAGTTCATCAGTAT | 169 |
|  | Ex2RT\_R | GAAAGAGTGATTCCCAGCAAATG |  |

**Table S2** Continued.

|  |  |  |  |
| --- | --- | --- | --- |
| **Genes** | **Primer designations** | **Primer sequences (5’→3’)** | **PCR product sizes [bp]** |
| **Real time PCR analyses** |
| *FOXL2* | Ex1RT\_F | TCAGCCTCAACGAGTGCTTCAT | 148 |
|  | Ex1RT\_R | GGTAGTTGCCCTTCTCGAACATG |  |
| *RSPO1* | Ex4RT\_F | GATGCAAGATTGAGCACTGCGA | 148 |
|  | Ex4RT\_R | GGGCTGCCGCACTCCAT |  |
| *SOX3* | Ex1RT\_F | GTGAACGGCTGGGCCAA | 112 |
|  | Ex1RT\_R | GCTGCTCATGGTCGCGG |  |
| *SOX9* | 800kb 5‘RT\_F | AAGGCAGTGTTCCAAACTGCTAG | 121 |
|  | 800kb 5‘RT\_R | TCTAAGGATCACAGTAGGAATCTCTCGT |  |
|  | 1kb 5‘RT\_F | CTTTCATTGATTCTCTGCGCC | 107 |
|  | 1kb 5‘RT\_R | CCGTTCCTCTGTAATAATCCGC |  |
|  | 0.9kb 5‘RT\_F | TGCTTCAAAAGACAATTTCTAAGCAC | 124 |
|  | 0.9kb 5‘RT\_R | GAATACTGCGAAGCCAGTTAAGTCC |  |
|  | 0.36kb 5’UTR RT\_F | GGAGCTTGAAACTGACCTGGAAC | 111 |
|  | 0.36kb 5’UTR RT\_R | CAATCGCCCCCGCTTTC |  |
|  | Ex1RT\_F | ATGAATCTCCTGGACCCCTTCATGAA | 128 |
|  | Ex1RT\_R | GTGTCGGAGCCGGAGCCCGAAGGGCA |  |
|  | IVS1 RT\_F | ggttcggcagaagcagcttt | 104 |
|  | IVS1 RT\_R | GGCCTTTGCTACTGGATATTTACCC |  |
|  | Ex2RT\_F | GCGTGCAGCACAAGAAGGAT | 142 |
|  | Ex2RT\_R | GCCTGCAGCGCCTTGAA |  |
|  | IVS2 RT\_F | gttaatcattctgtgccttatctcgg | 114 |
|  | IVS2 RT\_R | CTGCGCAGAGGCTCTGGTC |  |
|  | Ex3RT\_F | CTACATGAGCCCGGCGCAGCGGCCC | 116 |
|  | Ex3RT\_R | GCTGTGTGTACACCGGCTGTTCCCA |  |
|  | 1kb 3’UTR RT\_F | AGTGCCCCAGCACACTGC | 115 |
|  | 1kb 3’UTR RT\_R | GAAGGTTAACTGCTGGTGTTCAGAG |  |
|  | 1.5kb 3‘RT\_F | ATGTAGTGTATCACTGAGTCATTCGCA | 143 |
|  | 1.5kb 3‘RT\_R | TCCCCTCTCTCTCCAAGAGACA |  |
|  | 10kb 3‘RT\_F | CTGACTCAACACAGCATCGGA | 134 |
|  | 10kb 3‘RT\_R | TCCACACTCTGTGCCCAGG |  |
| *SOX10* | Ex2RT\_F | ATGGCAGAGGAGCAGGACCT | 189 |
|  | Ex2RT\_R | GCCGTCCTGCTGTTCCTTCT |  |
|  | Ex4RT\_F | TCCGACCACCCCAAAGACA | 144 |
|  | Ex4RT\_R | ATGTTGGACATTACCTCATGGCTG |  |
| *SRY* | Ex1RT\_F | GGTATGAGTGGAAAAGGCTTACAGAT | 159 |
|  | Ex1RT\_R | TTGAAGAGTCTGCAGGAAGCAAT |  |
| *WT1* | Ex1RT\_F | GTCATCCAGGATATTCCTGCTGTC | 148 |
|  | Ex1RT\_R | TGGAGTACTGCGGGTAGTTGTAGAG |  |
|  | Ex7RT\_F | AACCCGGAGAACCTCCACG | 139 |
|  | Ex7RT\_R | CGCTCGTTTTGGCTTCCG |  |
| *albumin* | Ex12RT\_F | AGTCACCAAATGCTGCACGG | 141 |
|  | Ex12RT\_R | GCATCGGGAAGGGTGCATAT |  |
| **Long-range PCR analyses** |
| *SOX9* | Ex3RT\_F | CTACATGAGCCCGGCGCAGCGGCCC | ≥1900 |
|  | 5‘-768\_R | GGACTTAACTGGCTTCGCAGTATTC |  |
|  | Ex3RT\_F | CTACATGAGCCCGGCGCAGCGGCCC | ≥2000 |
|  | 3’UTR+1504\_F | ATGTAGTGTATCACTGAGTCATTCGCA |  |
|  | 5‘-998\_F | CTTTCATTGATTCTCTGCGCC | 255# |
|  | 5‘-768\_R | GGACTTAACTGGCTTCGCAGTATTC |  |
|  | Ex3RT\_F | CTACATGAGCCCGGCGCAGCGGCCC | 1776 |
|  | 3‘UTR+1625\_R | TCCCCTCTCTCTCCAAGAGACA |  |

1 Pajares et al. 2007

2 Takahashi et al. 1998

\* Primers used for sequence analyses are tailed with M13 Tails (F: GTAAAACGACGGCCAGT; R: CAGGAAACAGCTATGAC)

# control systems for long range PCR