Supplementary Materials

Materials and Methods

1.1 Sample collection and sequencing

We collected a male *V. ferrilata* sample for genome sequencing, transcriptome sequencing and genome assembly from Gande County, Tibetan Autonomous Prefecture of Golog, Qinghai Province (33°541″ N, 99° 4854″ E). The white muscle from the legs and the blood were immediately placed into a dry ice kit to ensure the accuracy of sequencing. For RNA extraction, six organs (heart, liver, lung, gut, testis, and kidney) were cut into pieces and mixed with RNAlater. The samples were stored in an ultra-low temperature refrigerator at -80 ℃ for further use. All sample collection and experiments were approved by the Qinghai Forestry and Grassland Bureau and conformed to the guidelines established by the Ethics Committee for the Care and Use of Laboratory Animals of the Qufu Normal University (permit number: QFNU2018-013). See the supplementary appendix for relevant approval procedures.

The white muscle of *V. ferrilata* was ground into powder in liquid nitrogen, and sufficient high-quality DNA was extracted using the cetyltrimethylammonium bromide (CTAB) method (Stewart and Via, 1993). The obtained DNA was purified and enriched using the Blood & Cell Culture DNA Midi kit (Qiagen, Germany) and Oligo (dT) 0.5X magnetic beads. Then, DNA was detected and evaluated using 0.7% agarose gel pulse (Lonza) electrophoresis, Qubit (Thermo Fisher Scientific, USA), and Nanodrop (Thermo Fisher Scientific, USA). MGIEasy Universal DNA Library Prep Kit V1.0 （CAT#1000005250, MGI) was used to construct a paired‐end sequencing library following the standard protocol. The library was sequenced on the MGISEQ2000 platform according to the manufacturer’s instructions. We used the SageHL HMW library system (Sage Science, USA) to select about 8-10 μg of gDNAs (>50 kb) for each Nanopore library and processed them using the Ligation Sequencing 1D kit (SQK-LSK109, Oxford Nanopore Technologies, UK). Approximately 800 ng of DNA libraries were constructed and sequenced by Promethion (Oxford Nanopore Technologies, UK) at the Grandomics Genome Centre in Wuhan, China.

In the Hi-C experiment, chromatin was digested with 400 U of MboI restriction enzyme (NEB) after cross-linking. Then, the DNA ends were labeled with biotin, incubated for 45 min at 37 ℃, and the enzyme was inactivated in 20% SDS solution. T4 DNA ligase (NEB) was added to ligate the DNA, and the sample was incubated at 16 ℃ for 4-6 hours. After ligation, protease K was added for reverse cross-linking and incubated overnight at 65 ° C. DNA fragments were purified and dissolved in water. The non-ligated end was removed subsequently. The length of the purified DNA fragment was 300-500 BP, and the DNA ends were repaired. Finally, we isolated biotin-labeled DNA fragments from Dynabeads® M-280 Streptavidin (Life Technologies). The quality of the HI-C library was controlled and sequenced using an Illumina sequencer.

1.2 RNA extraction and sequencing

Six tissues (heart, liver, lung, gut, testis, and kidney) were ground into powder with TRIzol reagent (TIANGEN Biotech Co. Ltd, Beijing, China) on dry ice to extract total RNA. The integrity, concentration, and purity of the RNA were determined using an Agilent 2100 Bioanalyzer (Agilent Technologies, USA), 0.7% agarose gel pulse (Lonza) electrophoresis, and Nanodrop microspectrophotometer (Thermo Fisher Scientific, USA) respectively. Then, we enriched the mRNA using Oligo (dT) 0.5X magnetic beads. After enrichment, the mRNA were fragmented into short fragments by fragmentation buffer, and a cDNA strand was synthesized with random hexamers using mRNA as template. The buffer, dNTPs, DNA polymerase I and RNase H were added to synthesize two-stranded cDNA. Subsequently, AMPure XP Beads were used to purify double-stranded cDNA and the PCR products to obtain the final library. Finally, the quality of the library was controlled and sequenced using an Illumina sequencer.

1.3 Genome characteristics estimation

A DNA library of small fragments was constructed based on the characteristics of the genome using the genome-wide shotgun method (WGS). Whole-genome sequencing data were obtained through paired-end sequencing. Then, fastp 0.19.4 (https://github.com/OpenGene/fastp) software (-n 0 -f 5 -F 5 -t 5 -T 5) was used to filter the original data to obtain clean data. The filtering criteria included deleting the adapter sequence reads, intercepting the inaccurate sequencing bases at both ends of the reads, removing the reads containing N bases, and discarding the corresponding pair of reads when more than 20% of the basic mass fraction of the read was less than 20.

We used the k-mer analysis method to estimate genomic size and heterozygosity to estimate genomic characteristics from the read information obtained by sequencing. Different k-mer frequencies were obtained by processing reads with Jellyfish v0.20.0 (Marcais and Kingsford, 2011), and the genome size was estimated using the corresponding software (Find GSE, Genome scope). Kmerfreq uses its own algorithm to cut the K-mer (K-mer =19). The genome size was estimated using the K-mer frequency distribution (Luo et al., 2012).

1.4 De novo assembly, assessment and systeny analysis of the *Vulpes ferrilata* genome

After quality control, the filtered reads were used for pure third-generation assembly, and the genome was assembled using NextDenovo v2.0-beta.1 (READS_cutoff: 1 K, SEED_Cutoff: 25 K), which was mainly divided into three steps: first, the NextCorrect module was used to correct the original data and obtain a consistent sequence (CNS sequence) after the correction. Then, the NextGraph module was used for genome assembly of the corrected reads to obtain the preliminary assembly of the genome. Finally, the original third-generation and second-generation data were successively used for genome correction by Racon v1.3.1 and Nextpolish v1.0.5 (default third-generation data were corrected three times and second-generation data were corrected four times), and the polished genome was obtained. BUSCO assessment refers to the assessment of genomic integrity at the level of gene content

(Simão et al., 2015). For genome assembly evaluation, tBLASTn (Schffer et al., 2006) was compared with the corresponding BUSCO database sequences to identify candidate regions, Augustus software (Stanke et al., 2008) was used for gene structure prediction, and HMMER3 (Jaina et al., 2013) was used for comparison to assess its integrity. If the match length is within the expected range of the BUSCO profile match length, it is classified as "done." If they are found more than once, they are classified as "repeated." Matches that are only partially recovered are classified as "fragments" and those that are not are classified as "missing.” The genome was evaluated based on default parameters using BUSCO through the BUSCO single-copy gene database (mammalia_odb10). CEGMA (Genis et al., 2007) evaluation is based on a conserved protein family set (248 core genes) that exists in a large number of eukaryotes, and evaluates the assembled genome and the accuracy and integrity of the core genes within this genome. Using CEGMA v2 (default), the program uses information from the core genes of six model organisms to first identify candidate regions in the new genome using tBLASTn (Schffer et al., 2006). The system uses a profile of each core gene to ensure the reliability of the final prediction of the gene structure. Sequence consistency assessments aim to use high-quality second-generation sequencing data to assess the accuracy of third-generation sequencing data assembly results at the single-base level. The sources of single-base variation in third-generation sequencing data include:(1) real single-base variation, and (2) single base variation caused by sequencing error. When a single base variation caused by a sequencing error exists in the reference genome, it will be identified as a homozygous single-base variation when detecting single-base variations using second-generation sequencing data. Therefore, the final version of the genome assembled by the third-generation data can be used as the reference genome, single-base variation can be detected by the second-generation data, and the homozygous single-base variation rate can be used as the error rate of the assembly result. BWA 0.7.12-r1039, samtools v1.4, and bcftools v1.8.0 were used to detect the single base variation. The main process is divided into three steps: (1) The final genome assembled by three generations of data was used as the reference genome in BWA 0.7.12-r1039 analysis (Li et al., 2013). The software compared the filtered second-generation data (FASTP: -N 0 for second-generation data filtering) with the default parameters of MEM mode to the reference genome to obtain the comparison file; (2) SAMtools (Li et al., 2009) and bcftools (Danecek et al., 2017) were used to obtain the VCF file of the variation result; (3) the number of homozygous single base variations in VCF files were counted and the homozygous single base variation rate (number of homozygous single base variation/total number of genome bases) calculated. For systematic analysis, all protein sequences were aligned using blastp (Jacob et al., 2008). Synteny was then found using MCScanX (Wang et al., 2012). Finally, a circle figure was displayed using Circos (Krzywinski et al., 2009).

1.5 Chromosome assembly by Hi-C data

HIC Pro v2.8.1, a previous research achievement generated from the libraries, was used to control the quality of Hi-C raw data. Then, the low-quality sequences (quality scores < 20), adaptor sequences, and sequences shorter than 30 bp were filtered out

using fastp (-end-to-end -very-sensitive -L 30), and bowtie2 was used to map the clean paired-end reads to the draft assembled sequence to obtain the unique mapped paired-end reads (Chen et al., 2018; Langmead and Salzberg, 2012). Next, valid interactive paired reads from uniquely mapped paired-end reads were identified and retained by HIC-pro v2.8.1 to further analyze and filter invalid read pairs, including dangling-end, self-cycle, re-ligation, and dumped products. The obtained scaffolds were further aggregated and ordered, and the scaffolds were oriented to chromosomes using LACHESIS (https://github.com/shendurelab/LACHESIS). The following parameters were used for the analyses: CLUSTER_MIN_RE_SITES=100, CLUSTER_MAX_LINK_DENSITY=2.5, CLUSTER NONINFORMATIVE RATIO=1.4, ORDER MIN N RES IN TRUNK=60, ORDER MIN N RES IN SHREDS=60. Finally, the positioning errors were adjusted manually, which showed distinct patterns of discrete chromatin interactions (Burton et al., 2013).

1.6 Repeat annotation

Firstly, GMATA v2.2 (default) (Wang and Wang, 2016) and Tandem Repeats Finder 4.07b (2 7 7 80 10 50 500 -f -d -h -r) (Benson, 1999) software were used to identify the simple repeat sequences (SSRs) and recognize and annotate all tandem repeat elements (TRs) in the whole genome. Transposable elements (TEs) were identified by a combination of ab initio and homology-based methods in the *V. ferrilata* genome. Briefly, an ab inito repeat library was predicted by MITE-hunter (-n 20 -P 0.2 -c 3) (Han and Wessler, 2010) and Repeat Modeler version open-1.0.11 (-engine wublast) (Bedell et al., 2000) with default parameters. The obtained library was aligned using TEclass Repbase (Jurka et al., 2005) to classify the types of repetitive families. Repeat Masker 1.331 (nolow -no_is -gff -norna -engine abblast -lib lib) was also used to search for known and new TEs for further identification of repeats throughout the genome by mapping sequences against the de novo repeat library and Repbase TE library. The overlapping transposable elements belonging to the same repeat class were sorted and combined.

1.7 Gene prediction

Three independent methods: ab initio prediction, homology search, and transcriptome prediction, were used for gene prediction. We used GeMoMa v1.6.1 (Keilwagen et al., 2016) to conduct a homology search and obtain gene structure information by aligning homologous proteins from related species. For RNAseq-based gene prediction, STAR 2.7.3a (--outWigType bedGraph --outSAMtype BAM SortedByCoordinate --outSAMstrandField intronMotif) was used to align the filtered RNA-seq with the reference genome. PASA 2.3.3 (-c alignAssembly.config -C -R -g genome.fasta -T -u trans.fasta -t trans.clean.fasta -f fl.acc --CPU 10 --ALIGNERS gmap) (Haas et al., 2008) was used to predict transcripts composed of stringtie v1.3.4d and open reading frames (ORFs) after alignment. We used Augustus v3.3.1 (--gff3=on --hintsfile=hints.gff --extrinsicCfgFile=extrinsic.cfg --allow_hinted_splicesites=gcag,atac --min_intron_len=30 --softmasking=1) (Stanke et al., 2008) with the training set produced by RNA-seq reads to predict the ab initio gene. Then, EVidenceModeler v1.1.1 (--segmentSize 1000000 --overlapSize 100000) was used to generate a complete gene set. We used PASA 2.3.3 to compare the

transcripts to the genomes, and to predict the UTR and variable spliced regions of the original gene set based on transcriptome data. Finally, we used the TransposonPSI package (http://transposonpsi. sourceforge.net/) to remove the gene with TE and filter the miscoded genes (Urasaki et al., 2016).

1.8 Functional annotation of gene models

Gene functional information, motif, and domain were determined by comparison with public databases, such as Gene Ontology, KOG, KEGG, NR, and SWISSPORT. The GO terms were determined using the InterProScan 5.32-71.0 with default parameters (Zdobnov and Apweiler, 2001). We used BLASTP v2.7.1 (-max_target_seqs 1, -evalue 1e-5) to compare the integrated protein sequences with the other four public protein databases. The E cut-off value was 1e−05, while the hit result with the lowest E value was retained. The results of the five database searches were concatenated.

1.9 Annotation of non-coding RNAs (ncRNAs)

There are two strategies to obtain ncRNA: database retrieval and model prediction. The tRNAscan-SE v2.0 (--thread 4 -E -I) with eukaryote parameters was used to predict transfer RNAs (tRNAs) (Lowe and Eddy, 1997). We searched for microRNA, rRNA, small nuclear RNA, and small nucleolar RNA in the Rfam database using Infernal cmscan 5.32-71.0 (Nawrocki and Eddy, 2013; Griffiths-Jones, 2004). We used RNAmmer v1.2 (-S euk -m lsu,ssu,tsu -gff) to predict rRNAs and their subunits (Lagesen et al., 2007).

1.10 Identification of homologous and orthologous gene set

The homologous relationships between *V. ferrilata* and other assessed species (*Ailuropoda melanoleuca, Canis lupus, Enhydra lutris, Felis catus, Homo sapiens, Leptonychotes weddel, Mus musculus, Mustela putorius, Panthera tigris, Vulpes vulpes*) were confirmed by aligning the protein sequences using OrthMCL v2.0.9 (percentMatchCutoff=50 evalueExponentCutoff=-5) (Li and Stoeckert, 2003). Specifically, the prematurely terminated and miscoded genes of the protein sets collected from 10 sequenced species were filtered, and the longest transcript of each gene was retained. Then, the extracted protein sequences were aligned in pairs, and BLASTP 2.6.0+ was set to the E value threshold (≤ 1x10-5) to identify conserved homologous genes. OrthMCL was used to further identify homologous gene pairs, single-copy gene pairs, and paralogous intra-genome gene pairs. Finally, we identified the unique genes of *V. ferrilata* and the unclustered genes that exhibited no homologs as species-specific genes. The species-specific genes were further enriched, and their functions were annotated using the information on homologs in the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (http://www.geneontology.org/) databases.

1.11 Phylogenetic analyses

Based on the orthologous gene set, single-copy genes were used for molecular phylogenetic analysis. Mafft v7.313 (linsi) was used for multiple alignments of the coding sequences, and the ortholog group obtained from single-copy families (Katoh and Standley, 2013). Gblocks 0.91b (-t=p -b5=h) (Castresana et al., 2000) and RAXML 8.2.10 (-m PROTGAMMAAUTO -p 12345 -T 8 -f b) (Stamatakis et al., 2006) was used for low alignment sequence deletion and phylogenetic tree

construction. Gblocks (Castresana, 2000) and the GTRGAMMA substitution model of RAxML (Stamatakis et al., 2006) were used to eliminate the poorly aligned sequences and construct the phylogenetic tree respectively. The generated tree file was displayed with Figtree 1.4.4 and MEGA 10.1.8. Based on the phylogenetic tree, we used Real-time of Mega-CC to estimate the time of species differentiation by calculating the mean substitution rate of each branch (Kumar et al., 2012). We obtained six fossil calibration dates from the Timetree database (http://www.timet ree.org/) as time controls.

1.12 Gene family expansion and contraction analysis

The significant expansion or contraction of specific gene families is usually related to the adaptive differentiation of related species. The expansion and contraction of homologous gene families was detected by CAFE v4.2.1 (-p 0.05 -t 10 -r 10000) based on the results of Orthomcl, which uses birth and death processes to simulate gene gains and losses during phylogeny (Bie et al., 2006). For gene enrichment analysis of GO and KEGG pathways, the total gene dataset was used as the background, and the targeted genes searched the background and constructed the enriched pathway and GO terms using clusterProfiler (Yu et al., 2012).

1.13 Genes under positive selection

Genes with natural selection signatures can be identified by the ratio of the non-synonymous substitution rate (ka) to the synonymous substitution rate (ks) of protein-coding genes, according to the theory of neutral molecular evolution. Therefore, the average Ka/Ks values were calculated, and CodeML implemented in PAML Version 4.8 (Yang, 2007) was used to test the likely ratio of branching sites to determine the positive selection genes in the *V. ferrilata* genome. The genes with $p < 0.05$ were positive selection genes.

References

Bedell JA, Korf I, Gish W. 2000. MaskerAid: a performance enhancement to RepeatMasker. *Bioinformatics*, **16**(11): 1040-1041. doi: 10.1093/bioinformatics/btt509

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, **27**(2): 573-580. doi: 10.1093/nar/27.2.573

Bie TD, Cristianini N, Demuth JP, Hahn AMW. 2006. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*, **22**(10): 1269-1271. doi: 10.1093/bioinformatics/btl097

Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J, et al. 2013. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature Biotechnology*, **31**(12): 1119-1125. doi: 10.1038/nbt.2727

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*, **17**(4): 540-552. doi: 10.1093/oxfordjournals.molbev.a026334

Chen S, Zhou Y, Chen Y. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics,* **34**(17): i884-i890. doi: 10.1101/274100

Danecek P, Mccarthy SA. 2017. BCFtools/csq: Haplotype-aware variant consequences. doi: 10.1101/090811

Genis P, Keith B, Ian K. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic

genomes. *Bioinformatics*, **23**(9): 1061-1067. doi: 10.1093/bioinformatics/btm071

Griffiths-Jones S. 2004. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Research*, **33**(Database issue): D121-D124. doi: 10.1093/nar/gki081

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome biology*, **9**(1): R7. doi: 10.1186/gb-2008-9-1-r7

Han Y, Wessler SR. 2010. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Research*, **38**(22): e199. doi: 10.1093/nar/gkq862

Jacob A, Lancaster J, Buhler J, Harris B, Chamberlain R.D. 2008. Mercury blastp: accelerating protein sequence alignment. *Acm Transactions on Reconfigurable Technology & Systems*, **1**(2), 1-44. doi: 10.1145/1371579.1371581

Jaina M, Finn RD, Eddy SR, Alex B, Marco P. 2013. Challenges in homology search: hmmer3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research*, **41**(12), e121-e121. doi: 10.1093/nar/gkt263

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J, et al. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, **110**(1-4): 462-467. doi: 10.1159/000084979

Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, **30**(4): 772-780. doi: 10.1093/molbev/mst010

Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F, et al. 2016. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Research*, **44**(9): e89. doi: 10.1093/nar/gkw092

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D. et al. 2009. Circos: an information aesthetic for comparative genomics. *Genome Research*, **19**(9):1639-45. doi: 10.1101/gr.092759.109

Kumar S, Stecher G, Peterson D, Tamura K. 2012. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics*, **28**(20): 2685-2686. doi: 10.1093/nar/gkw092

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**(4): 357–359. doi: 10.1038/nmeth.1923

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 2009. The sequence alignment/map format and samtools. *Bioinformatics*, **25**(16), 2078-2079. doi: 10.1093/bioinformatics/btp352

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with bwa-mem. arXiv e-prints.

Li L, Stoeckert C. 2003. OrthoMCL: Identification of ortholog groups for Eukaryotic genomes. *Genome Research*, **13**(1088-9051/03): 2178-2189. doi: 10.1101/gr.1224503

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, **25**(5): 955-964. doi: 10.1093/nar/25.5.955

Luo RB, Liu BH, Xie YL, Li ZY, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, **1**(18). doi: 10.1186/2047-217X-1-18

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences

of k-mers. *Bioinformatics*, **27**(6): 764 – 770. doi: 10.1093/bioinformatics/btr011

Nawrocki E, Eddy S. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, **29**(22): 2933-2935. doi: 10.1093/bioinformatics/btt509

Schffer AA, Richa A, Yu YK, Michael GE, Altschul SF. 2006. Composition-based statistics and translated nucleotide searches: improving the tblastn module of blast. *BMC Biology*, **4**(1):41. doi: 10.1186/1741-7007-4-41

Simão FA, Waterhouse RM, Panagiotis I, Kriventseva EV, Zdobnov EM. 2015. Busco: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics,* **(19)**3210-3212. doi: 10.1093/bioinformatics/btv351

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, **24**(5): 637-644. doi: 10.1093/bioinformatics/btn013

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**(21): 2688-2690. doi: 10.1093/bioinformatics/btl446

Stewart C, Via LE. 1993. A rapid ctab dna isolation technique useful for rapd fingerprinting and other pcr applications. *Biotechniques*, **14**(5): 748-50.

Wang X, Wang L. 2016. GMATA: An integrated software package for genome-scale SSR mining, marker development and viewing. *Frontiers in Plant Science*, **13**(7): 1350. doi: 10.3389/fpls.2016.01350

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. 2012. Mcscanx: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, **40**(7), e49-e49. doi: 10.1093/nar/gkr1293

Urasaki N, Takagi H, Natsume S, Uemura A, Taniai N, Miyagi N, et al. 2016. Draft genome sequence of bitter gourd (Momordica charantia), a vegetable and medicinal plant in tropical and subtropical regions. *DNA Research*, **0**(0): 1-8. doi: 10.1093/dnares/dsw047

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**(8): 1586-1591. doi: 10.1093/molbev/msm088

Yu G, Wang LG, Han Y, He QY. 2012. Clusterprofiler: an r package for comparing biological themes among gene clusters. *Omics-a Journal of Integrative Biology*, **16**(5), 284-287. doi: 10.1089/omi.2011.0118

Zdobnov EM, Apweiler R. 2001. InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, **17**(9): 847-848. doi: 10.1093/bioinformatics/17.9.84

Supplementary Table S1. Sequencing statistics generated by different platforms for *Vulpes ferrilata* genome assembly

Platform	Library size	Raw data (Gb)	Clean data (Gb)	Coverage (X)
MGI	400 bp	155	145	58.98
Nanopore	20 kb	-	253	102.73
Hi-C	350 bp	264	252	112.8
Toal	-	-	653	274.51

Supplementary Table S2. Total pass reads information statistic of ONT sequencing platform

Library ID	Total pass reads bases	Mean length	N50 length
20200409-NPL1962-P5-PAE64028	77,841,186,926	16,465	22,793
20200410-NPL1962-P1-PAE66557	68,860,710,913	16,783	23,374
20200411-NPL1962-P5-PAE67654	59,657,248,207	16,620	23,466
20200413-NPL1962-P5-PAE68858	46,649,723,563	15,459	22,303
Total	253,008,869,609	16,389	23,002

Supplementary Table S3. Statistical analysis of transcriptome sequencing results of six organs

Sample	Total Reads	Total Bases (bp)	Clean Reads	Clean Bases (bp)	Q20 Rate(%)	Q30 Rate(%)	GC(%)
Gut	47,301,306	7,095,195,900	46,908,360	7,020,363,082	96.91	91.51	46.55
Heart	70,892,810	10,633,921,500	70,408,436	10,512,563,558	98.64	95.57	47.57
Kidney	48,578,428	7,286,764,200	48,165,658	7,209,562,145	96.94	91.58	47.61
Liver	45,734,350	6,860,152,500	45,352,874	6,761,981,620	97.25	92.32	48.29
Lung	47,522,488	7,128,373,200	47,053,648	7,027,857,700	97.14	92.14	47.69
Spleen	50,878,684	7,631,802,600	50,347,522	7,509,082,471	97.04	91.91	47.77

Supplementary Table S4. Statistics of assembly results

Stat type	Preliminary Assembly		Polish genome	
	Contig length	Contig number	Contig length	Contig number
N50	61,004,393 bp	14	61,591,022	14
Longest	130,478,765 bp	1	131,792,797	1
Total	2,354,730,612	359	2,379,644,749	359
Length>=1kb	2,354,730,612	359	2,379,644,749	359
Length>=2kb	2,354,730,612	359	2,379,644,749	359
Length>=5kb	2,354,730,612	359	2,379,644,749	359

Supplementary Table S5. BUSCO Predictive Statistics

Type	Number	Percent (%)
Complete BUSCOs (C)	8,558	92.76
Complete and single-copy BUSCOs (S)	8,477	91.88
Complete and duplicated BUSCOs (D)	81	0.88
Fragmented BUSCOs (F)	159	1.72
Missing BUSCOs (M)	509	5.52
Total BUSCO groups searched	9,226	100.00

Supplementary Table S6. Genome single base accuracy statistics

Depth(X)	Hetero SNP	Hetero Indel	Homo SNP	Error rate by Homo SNP(%)	Homo Indel	Error rate by Homo Indel(%)	Error rate by homo variants(%)	Accuracy of genome(%)
depth>=1x	2,099,929	752,271	12,016	0.000505	35,212	0.00148	0.001985	99.998015
depth>=5x	2,098,745	749,377	8,282	0.000348	29,587	0.001243	0.001591	99.998409
depth>=10x	2,093,224	734,593	6,508	0.000273	22,588	0.000949	0.001223	99.998777

Supplementary Table S7. Chromosome and reference genome corresponding chromosome statistical results

Chr	Size(bp)	Contig Num
LG01	203,316,443	6
LG02	166,046,677	20
LG03	157,658,846	7
LG04	157,089,576	7
LG05	136,349,340	3
LG06	135,942,581	22
LG07	134,372,329	3
LG08	133,959,777	8
LG09	126,600,798	9
LG10	125,059,525	28
LG11	122,981,691	9
LG12	122,878,232	5
LG13	122,537,276	3
LG14	119,274,077	6
LG15	118,459,051	9
LG16	113,145,118	5
LG17	80,532,236	4
LG18	64,503,012	3
Total	2,340,706,585	157

Supplementary Table S8. Hi-C assembly result statistics

Stat Type	Scaffold Length	Scaffold Number	Contig Length	Contig Number	Gap Length	Gap Number
N50	133,960,477	8	53,181,253	17	100	70
N60	125,062,225	10	42,800,000	22	100	84
N70	122,878,632	12	35,421,494	28	100	98
N80	119,274,577	14	25,832,893	36	100	112
N90	113,145,518	16	17,393,552	47	100	126
Longest	203,316,943	1	116,451,844	1	100	1
Total	2,379,658,649	234	2,379,644,749	373	13,900	139
Length>=1kb	2,379,658,649	234	2,379,644,749	373	0	0
Length>=2kb	2,379,658,649	234	2,379,644,749	373	0	0
Length>=5kb	2,379,658,649	234	2,379,644,749	373	0	0

Supplementary Table S9. Statistical results of TE repeat sequences

Class	Order	Super family	Number of elements	Length of sequence(bp)	Percentage of sequence(%)
Class I			2726505	702457281	29.52
	LTR		711910	148044339	6.22
		ERVL-MaLR	181392	47419317	1.99
		Unknown	218195	28274984	1.19
		ERV1	111836	28871158	1.21
		Gypsy	43818	3989130	0.17
		ERVL	124921	37980651	1.6
		Other	31748	1509099	0.06
	SINE		860533	115207788	4.84
		B2	337234	46190669	1.94
		Unknown	228526	32344796	1.36
		MIR	246348	34717533	1.46
		Other	48425	1954790	0.08
	LINE		1154062	439205154	18.46
		L2	201606	37380859	1.57
		L1	874870	390958759	16.43
		CR1	25815	3955075	0.17
		Unknown	7284	2914002	0.12
		Other	44487	3996459	0.17
Class II			682507	78353408	3.29
	DNA		665015	77371204	3.25
		hAT-Charlie	190224	31863352	1.34
		hAT-Tip100	53841	8505619	0.36
		Unknown	40663	2447469	0.1
		TcMar-Tigger	59814	15052314	0.63
		Other	320473	19502450	0.82
	RC		17492	982204	0.04
		Other	17492	982204	0.04
Total TEs			3409012	780810689	32.81
Tandem Repeats			475481	18348581	0.77
	tandem_repeat		242635	15340903	0.64
	SSR		232846	3007678	0.13
Other			613241	90970079	3.82
Unknown			20339	3901317	0.16
Simple repeats			25742	4587768	0.19
Low complexity			1446	242363	0.01

TotalRepeats	4545261	898860797	37.77

Supplementary Table S10. Protein-coding gene annotation statistics

Gene set	Total number of genes	Average gene length(bp)	Average CDS length(bp)	Average exons number per gene	Average exon length(bp)	Average intron length(bp)
V. ferrilata	21,715	32,053.47	1,330.64	7.53	176.75	4,706.1

Supplementary Table S11. Gene functional annotation statistics based on gene functional annotation results from different databases

Type		Number	Percent (%)
Annotation	Swissprot	18,358	84.54
	KEGG	13,661	62.91
	KOG	12,520	57.66
	GO	12,495	57.54
	NR	18,580	85.56
Total	Annotated	18,776	86.47
	Gene	21,715	-

Supplementary Table S12. Non-coding RNA annotation statistics

Type		Copy Number	Average Length(bp)	Total Length(bp)	Percentage of sequence(%)
rRNA (295)	18S	17	1,884.12	32,030	0.0013
	28S	17	6,951.35	118,173	0.005
	5.8S	15	152.6	2,289	0.0001
	5S	246	111.67	27,471	0.0012
miRNA (3,788)	snRNA	560	109.57	61,358	0.0026
	miRNA	1,068	78.11	83,424	0.0035
	spliceosomal	1,600	118.01	188,818	0.0079
	other	560	229.32	128,417	0.0054
Regulatory	cis-regulatory elements	585	65.85	38,524	0.0016
tRNA	tRNA	143,320	82.76	11,861,506	0.4985

Supplementary Table S13. 67 significantly over-represented KEGG pathways

ID	Description	Gene Ratio	Bg Ratio	pvalue	padj	Count
map04966	Collecting duct acid secretion	44/135	76/7235	6.28E-59	5.65E-57	44
map04145	Phagosome	56/135	237/7235	4.36E-49	1.96E-47	56
map04721	Synaptic vesicle cycle	44/135	132/7235	2.62E-45	7.85E-44	44
map00190	Oxidative phosphorylation	44/135	195/7235	5.88E-37	1.32E-35	44
map04150	mTOR signaling pathway	44/135	221/7235	2.02E-34	3.64E-33	44
map04213	Longevity regulating pathway - multiple species	19/135	93/7235	3.49E-15	5.23E-14	19
map00710	Carbon fixation in photosynthetic organisms	21/135	136/7235	4.11E-14	5.29E-13	21
map00010	Glycolysis / Gluconeogenesis	21/135	178/7235	9.15E-12	9.15E-11	21
map01230	Biosynthesis of amino acids	21/135	178/7235	9.15E-12	9.15E-11	21
map04623	Cytosolic DNA-sensing pathway	12/135	57/7235	3.77E-10	3.39E-09	12
map04915	Estrogen signaling pathway	19/135	176/7235	4.63E-10	3.79E-09	19
map04066	HIF-1 signaling pathway	21/135	224/7235	7.28E-10	5.46E-09	21
map04612	Antigen processing and presentation	13/135	80/7235	2.02E-09	1.40E-08	13
map04622	RIG-I-like receptor signaling pathway	12/135	66/7235	2.27E-09	1.46E-08	12
map01200	Carbon metabolism	21/135	252/7235	6.25E-09	3.75E-08	21
map04611	Platelet activation	15/135	144/7235	5.85E-08	3.19E-07	15
map04217	Necroptosis	17/135	188/7235	6.02E-08	3.19E-07	17
map04913	Ovarian steroidogenesis	10/135	59/7235	1.06E-07	5.29E-07	10
map04620	Toll-like receptor signaling pathway	12/135	96/7235	1.79E-07	8.48E-07	12

map04650	Natural killer cell mediated cytotoxicity	12/135	98/7235	2.26E-07	1.02E-06	12
map01120	Microbial metabolism in diverse environments	21/135	318/7235	3.54E-07	1.52E-06	21
map04918	Thyroid hormone synthesis	11/135	85/7235	4.18E-07	1.71E-06	11
map04923	Regulation of lipolysis in adipocytes	10/135	75/7235	1.08E-06	4.24E-06	10
map04540	Gap junction	13/135	142/7235	2.08E-06	7.81E-06	13
map04921	Oxytocin signaling pathway	15/135	190/7235	2.18E-06	7.83E-06	15
map04720	Long-term potentiation	10/135	82/7235	2.50E-06	8.66E-06	10
map04927	Cortisol synthesis and secretion	10/135	83/7235	2.80E-06	9.33E-06	10
map04976	Bile secretion	10/135	84/7235	3.13E-06	1.01E-05	10
map01130	Biosynthesis of antibiotics	21/135	377/7235	5.58E-06	1.73E-05	21
map04141	Protein processing in endoplasmic reticulum	15/135	206/7235	5.94E-06	1.78E-05	15
map04971	Gastric acid secretion	10/135	91/7235	6.52E-06	1.89E-05	10
map04970	Salivary secretion	10/135	95/7235	9.62E-06	2.71E-05	10
map04714	Thermogenesis	17/135	283/7235	1.81E-05	4.94E-05	17
map04727	GABAergic synapse	10/135	103/7235	1.98E-05	5.09E-05	10
map04972	Pancreatic secretion	10/135	103/7235	1.98E-05	5.09E-05	10
map04630	Jak-STAT signaling pathway	12/135	150/7235	2.10E-05	5.24E-05	12
map04911	Insulin secretion	10/135	104/7235	2.15E-05	5.24E-05	10
map04270	Vascular smooth muscle contraction	12/135	151/7235	2.24E-05	5.31E-05	12
map04912	GnRH signaling pathway	10/135	106/7235	2.55E-05	5.88E-05	10
map04015	Rap1 signaling pathway	15/135	235/7235	2.89E-05	6.50E-05	15

map04621	NOD-like receptor signaling pathway	12/135	160/7235	4.00E-05	8.77E-05	12
map04926	Relaxin signaling pathway	11/135	136/7235	4.24E-05	9.09E-05	11
map04713	Circadian entrainment	10/135	113/7235	4.45E-05	9.09E-05	10
map04925	Aldosterone synthesis and secretion	10/135	113/7235	4.45E-05	9.09E-05	10
map04211	Longevity regulating pathway	10/135	114/7235	4.80E-05	9.59E-05	10
map04750	Inflammatory mediator regulation of TRP channels	10/135	115/7235	5.17E-05	0.000101176	10
map04914	Progesterone-mediated oocyte maturation	10/135	118/7235	6.45E-05	0.000123468	10
map04916	Melanogenesis	10/135	119/7235	6.93E-05	0.000129918	10
map04725	Cholinergic synapse	10/135	124/7235	9.82E-05	0.000180358	10
map04210	Apoptosis	12/135	185/7235	0.000162641	0.000287117	12
map04060	Cytokine-cytokine receptor interaction	14/135	243/7235	0.00016 27	0.000287117	14
map04371	Apelin signaling pathway	11/135	159/7235	0.00017 4818	0.000293575	11
map04724	Glutamatergic synapse	10/135	133/7235	0.00017 6145	0.000293575	10
map04928	Parathyroid hormone synthesis, secretion and action	10/135	133/7235	0.00017 6145	0.000293575	10
map04261	Adrenergic signaling in cardiomyocytes	11/135	170/7235	0.00031 3803	0.000513496	11
map04114	Oocyte meiosis	10/135	152/7235	0.00051 8554	0.000833 39	10
map04723	Retrograde endocannabinoid signaling	10/135	157/7235	0.00066 8933	0.001056211	10
map04745	Phototransduction - fly	5/135	43/7235	0.00114 3916	0.001775042	5
map00230	Purine metabolism	10/135	172/7235	0.00134 9948	0.002059242	10
map04072	Phospholipase D signaling pathway	10/135	174/7235	0.00147 3004	0.002209506	10
map0	cGMP-PKG signaling	10/135	185/7235	0.00232	0.003428	10

4022	pathway			3658	348	
map04062	Chemokine signaling pathway	10/135	189/7235	0.002716862	0.003943832	10
map03040	Spliceosome	9/135	170/7235	0.004388317	0.006269024	9
map04020	Calcium signaling pathway	10/135	208/7235	0.005373021	0.007555811	10
map04024	cAMP signaling pathway	10/135	233/7235	0.011571186	0.016021642	10
map04391	Hippo signaling pathway - fly	5/135	87/7235	0.022930896	0.031269404	5
map04520	Adherens junction	5/135	90/7235	0.026105047	0.035066481	5

Supplementary Table S14. 32 significantly enriched GO-terms

ID	Description	GO Class	Gene Ratio	Bg Ratio	pvalue	padj	Count
GO:0033178	proton-transporting two-sector ATPase complex, catalytic domain	CC	44/285	51/13283	1.28E-67	7.55E-66	44
GO:0046961	proton-transporting ATPase activity, rotational mechanism	MF	44/285	56/13283	5.64E-64	1.66E-62	44
GO:0015991	ATP hydrolysis coupled proton transport	BP	44/285	79/13283	2.18E-53	4.28E-52	44
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	BP	30/285	92/13283	8.71E-28	1.28E-26	30
GO:0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	MF	33/285	157/13283	1.17E-23	1.38E-22	33
GO:0005126	cytokine receptor binding	MF	17/285	26/13283	7.06E-23	6.94E-22	17
GO:0006952	defense response	BP	17/285	30/13283	2.51E-21	2.11E-20	17
GO:0022857	transmembrane transporter activity	MF	23/285	137/13283	1.48E-14	1.09E-13	23
GO:0006006	glucose metabolic process	BP	9/285	13/13283	5.64E-13	3.70E-12	9
GO:0008641	ubiquitin-like modifier activating enzyme activity	MF	9/285	18/13283	3.49E-11	2.06E-10	9
GO:0003746	translation elongation factor activity	MF	9/285	22/13283	3.31E-10	1.78E-09	9
GO:0006414	translational elongation	BP	9/285	24/13283	8.38E-10	4.12E-09	9
GO:0050661	NADP binding	MF	9/285	27/13283	2.84E-09	1.29E-08	9
GO:0000786	nucleosome	CC	15/285	100/13283	3.21E-09	1.35E-08	15

GO:0051287	NAD binding	MF	9/285	29/13283	5.84E-09	2.30E-08	9
GO:0046982	protein heterodimerization activity	MF	15/285	123/13283	5.71E-08	2.11E-07	15
GO:0006879	cellular iron ion homeostasis	BP	7/285	24/13283	4.92E-07	1.61E-06	7
GO:0008199	ferric iron binding	MF	7/285	24/13283	4.92E-07	1.61E-06	7
GO:0006464	cellular protein modification process	BP	7/285	30/13283	2.59E-06	8.05E-06	7
GO:0009190	cyclic nucleotide biosynthetic process	BP	7/285	32/13283	4.13E-06	1.16E-05	7
GO:0016849	phosphorus-oxygen lyase activity	MF	7/285	32/13283	4.13E-06	1.16E-05	7
GO:0036459	thiol-dependent ubiquitinyl hydrolase activity	MF	8/285	59/13283	3.49E-05	9.37E-05	8
GO:0055085	transmembrane transport	BP	23/285	437/13283	6.96E-05	0.0001 7495	23
GO:0016579	protein deubiquitination	BP	8/285	65/13283	7.12E-05	0.0001 7495	8
GO:0005576	extracellular region	CC	17/285	270/13283	7.45E-05	0.0001 75727	17
GO:0005085	guanyl-nucleotide exchange factor activity	MF	5/285	40/13283	0.0015 61931	0.0035 44382	5
GO:0008234	cysteine-type peptidase activity	MF	4/285	37/13283	0.0078 53378	0.0171 61085	4
GO:0007264	small GTPase mediated signal transduction	BP	5/285	60/13283	0.0091 65502	0.0193 13022	5
GO:0007165	signal transduction	BP	15/285	357/13283	0.0102 92272	0.0209 3945	15
GO:0051082	unfolded protein binding	MF	5/285	67/13283	0.0143 6862	0.0282 58286	5
GO:0006457	protein folding	BP	5/285	74/13283	0.0212 5821	0.0404 59173	5
GO:0016567	protein ubiquitination	BP	3/285	30/13283	0.0258 84015	0.0477 23652	3

Supplementary Table S15. 175 positively selected genes (PSGs) in *Vulpes ferrilata*

Gene_ID	E value	Function
MRPL14	5.72E-86	39S ribosomal protein L14, mitochondrial OS=Bos taurus OX=9913 GN=MRPL14 PE=1 SV=1
CEP135	0	Centrosomal protein of 135 kDa OS=Homo sapiens OX=9606 GN=CEP135 PE=1 SV=2
Slc12a6	8.41E-41	Solute carrier family 12 member 6 OS=Mus musculus OX=10090 GN=Slc12a6 PE=1 SV=2
SLC16A6	0	Monocarboxylate transporter 7 OS=Homo sapiens OX=9606 GN=SLC16A6 PE=1 SV=2
Frmd5	0	FERM domain-containing protein 5 OS=Mus musculus OX=10090 GN=Frmd5 PE=1 SV=1
COL28A1	0	Collagen alpha-1(XXVIII) chain OS=Homo sapiens OX=9606 GN=COL28A1 PE=2 SV=2
Ccdc189	5.57E-139	Coiled-coil domain-containing protein 189 OS=Mus musculus OX=10090 GN=Ccdc189 PE=1 SV=2
COP1	0	E3 ubiquitin-protein ligase COP1 OS=Homo sapiens OX=9606 GN=COP1 PE=1 SV=1
PGF	5.99E-64	Placenta growth factor OS=Homo sapiens OX=9606 GN=PGF PE=1 SV=2
PRDM8	0	PR domain zinc finger protein 8 OS=Homo sapiens OX=9606 GN=PRDM8 PE=1 SV=3
CNST	0	Consortin OS=Homo sapiens OX=9606 GN=CNST PE=1 SV=3
MOG	2.02E-103	Myelin-oligodendrocyte glycoprotein OS=Bos taurus OX=9913 GN=MOG PE=1 SV=1
DNALI1	7.80E-135	Axonemal dynein light intermediate polypeptide 1 OS=Homo sapiens OX=9606 GN=DNALI1 PE=1 SV=2
FGF17	6.22E-132	Fibroblast growth factor 17 OS=Homo sapiens OX=9606 GN=FGF17 PE=1 SV=1
NDUFV2	2.74E-168	NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial OS=Pongo pygmaeus OX=9600 GN=NDUFV2 PE=2 SV=1
Abhd18	0	Protein ABHD18 OS=Rattus norvegicus OX=10116 GN=Abhd18 PE=2 SV=2
GPR156	0	Probable G-protein coupled receptor 156 OS=Homo sapiens OX=9606 GN=GPR156 PE=2 SV=2
P2RX2	0	P2X purinoceptor 2 OS=Homo sapiens OX=9606 GN=P2RX2 PE=1 SV=1
ITPRID1	0	Protein ITPRID1 OS=Homo sapiens OX=9606 GN=ITPRID1 PE=2 SV=2
NUP37	0	Nucleoporin Nup37 OS=Homo sapiens OX=9606 GN=NUP37 PE=1 SV=1
NR1H3	0	Oxysterols receptor LXR-alpha OS=Bos taurus OX=9913

		GN=NR1H3 PE=2 SV=1
ARPP21	0	cAMP-regulated phosphoprotein 21 OS=Homo sapiens OX=9606 GN=ARPP21 PE=1 SV=2
TDRKH	0	Tudor and KH domain-containing protein OS=Homo sapiens OX=9606 GN=TDRKH PE=1 SV=2
SYCP2	0	Synaptonemal complex protein 2 OS=Homo sapiens OX=9606 GN=SYCP2 PE=2 SV=2
LBX2	9.78E-69	Transcription factor LBX2 OS=Homo sapiens OX=9606 GN=LBX2 PE=2 SV=1
PNOC	5.26E-93	Prepronociceptin OS=Bos taurus OX=9913 GN=PNOC PE=2 SV=1
RHNO1	4.31E-91	RAD9, HUS1, RAD1-interacting nuclear orphan protein 1 OS=Homo sapiens OX=9606 GN=RHNO1 PE=1 SV=1
RBM45	0	RNA-binding protein 45 OS=Homo sapiens OX=9606 GN=RBM45 PE=1 SV=1
CCDC190	9.09E-87	Coiled-coil domain-containing protein 190 OS=Bos taurus OX=9913 GN=CCDC190 PE=2 SV=1
Pianp	7.50E-119	PILR alpha-associated neural protein OS=Mus musculus OX=10090 GN=Pianp PE=1 SV=1
INHBC	0	Inhibin beta C chain OS=Homo sapiens OX=9606 GN=INHBC PE=2 SV=1
CDC42SE1	2.46E-38	CDC42 small effector protein 1 OS=Homo sapiens OX=9606 GN=CDC42SE1 PE=1 SV=1
MTMR11	0	Myotubularin-related protein 11 OS=Homo sapiens OX=9606 GN=MTMR11 PE=2 SV=2
TCTE1	0	Dynein regulatory complex subunit 5 OS=Macaca fascicularis OX=9541 GN=TCTE1 PE=2 SV=1
Rbp3	0	Retinol-binding protein 3 OS=Mus musculus OX=10090 GN=Rbp3 PE=2 SV=3
PLPP5	7.08E-136	Phospholipid phosphatase 5 OS=Homo sapiens OX=9606 GN=PLPP5 PE=1 SV=2
Cdcp2	0	CUB domain-containing protein 2 OS=Mus musculus OX=10090 GN=Cdcp2 PE=2 SV=1
STIL	0	SCL-interrupting locus protein OS=Homo sapiens OX=9606 GN=STIL PE=1 SV=2
MAMSTR	9.35E-176	MEF2-activating motif and SAP domain-containing transcriptional regulator OS=Bos taurus OX=9913 GN=MAMSTR PE=2 SV=2
NLRX1	0	NLR family member X1 OS=Homo sapiens OX=9606 GN=NLRX1 PE=1 SV=1
POGZ	0	Pogo transposable element with ZNF domain OS=Homo sapiens OX=9606 GN=POGZ PE=1 SV=2
RWDD3	2.38E-122	RWD domain-containing protein 3 OS=Pongo abelii OX=9601 GN=RWDD3 PE=2 SV=1

PRRC2C	0	Protein PRRC2C OS=Homo sapiens OX=9606 GN=PRRC2C PE=1 SV=4
AK5	0	Adenylate kinase isoenzyme 5 OS=Homo sapiens OX=9606 GN=AK5 PE=1 SV=2
PUS7L	0	Pseudouridylate synthase 7 homolog-like protein OS=Homo sapiens OX=9606 GN=PUS7L PE=1 SV=1
Glp1r	0	Glucagon-like peptide 1 receptor OS=Rattus norvegicus OX=10116 GN=Glp1r PE=2 SV=1
Mbnl2	0	Muscleblind-like protein 2 OS=Rattus norvegicus OX=10116 GN=Mbnl2 PE=1 SV=1
DLG5	0	Disks large homolog 5 OS=Homo sapiens OX=9606 GN=DLG5 PE=1 SV=4
PSD4	0	PH and SEC7 domain-containing protein 4 OS=Homo sapiens OX=9606 GN=PSD4 PE=1 SV=2
Cmtm5	2.79E-88	CKLF-like MARVEL transmembrane domain-containing protein 5 OS=Mus musculus OX=10090 GN=Cmtm5 PE=1 SV=1
GPC5	0	Glypican-5 OS=Homo sapiens OX=9606 GN=GPC5 PE=2 SV=1
MAX	6.82E-105	Protein max OS=Homo sapiens OX=9606 GN=MAX PE=1 SV=1
SNTG1	0	Gamma-1-syntrophin OS=Homo sapiens OX=9606 GN=SNTG1 PE=1 SV=1
Map3k7cl	2.04E-69	MAP3K7 C-terminal-like protein OS=Mus musculus OX=10090 GN=Map3k7cl PE=2 SV=1
CHEK1	0	Serine/threonine-protein kinase Chk1 OS=Homo sapiens OX=9606 GN=CHEK1 PE=1 SV=2
SYNC	0	Syncoilin OS=Homo sapiens OX=9606 GN=SYNC PE=1 SV=3
Gtsf1	6.19E-115	Gametocyte-specific factor 1 OS=Mus musculus OX=10090 GN=Gtsf1 PE=1 SV=1
CUSTOS	5.47E-91	Protein CUSTOS OS=Bos taurus OX=9913 GN=CUSTOS PE=2 SV=1
XKR4	0	XK-related protein 4 OS=Homo sapiens OX=9606 GN=XKR4 PE=2 SV=1
Kcna7	0	Potassium voltage-gated channel subfamily A member 7 OS=Mus musculus OX=10090 GN=Kcna7 PE=2 SV=2
NPAS4	0	Neuronal PAS domain-containing protein 4 OS=Homo sapiens OX=9606 GN=NPAS4 PE=1 SV=1
MTNR1A	0	Melatonin receptor type 1A OS=Homo sapiens OX=9606 GN=MTNR1A PE=1 SV=1
BMP3	0	Bone morphogenetic protein 3 OS=Homo sapiens OX=9606 GN=BMP3 PE=1 SV=2
ASTL	0	Astacin-like metalloendopeptidase OS=Homo sapiens

		OX=9606 GN=ASTL PE=1 SV=4
RGSL1	0	Regulator of G-protein signaling protein-like OS=Homo sapiens OX=9606 GN=RGSL1 PE=2 SV=1
LGALS4	0	Galectin-4 OS=Homo sapiens OX=9606 GN=LGALS4 PE=1 SV=1
OSM	2.06E-67	Oncostatin-M OS=Homo sapiens OX=9606 GN=OSM PE=1 SV=2
MYCBPAP	0	MYCBP-associated protein OS=Homo sapiens OX=9606 GN=MYCBPAP PE=1 SV=2
UQCRB	5.43E-32	Cytochrome b-c1 complex subunit 7 OS=Bos taurus OX=9913 GN=UQCRB PE=1 SV=3
CD40LG	3.64E-165	CD40 ligand OS=Canis lupus familiaris OX=9615 GN=CD40LG PE=2 SV=1
TRMT10C	0	tRNA methyltransferase 10 homolog C OS=Bos taurus OX=9913 GN=TRMT10C PE=2 SV=2
METTL5	6.43E-138	Methyltransferase-like protein 5 OS=Homo sapiens OX=9606 GN=METTL5 PE=1 SV=1
TMEM94	0	Transmembrane protein 94 OS=Homo sapiens OX=9606 GN=TMEM94 PE=1 SV=1
DBX2	3.64E-152	Homeobox protein DBX2 OS=Homo sapiens OX=9606 GN=DBX2 PE=2 SV=2
Ceacam16	0	Carcinoembryonic antigen-related cell adhesion molecule 16 OS=Mus musculus OX=10090 GN=Ceacam16 PE=1 SV=1
NDRG4	0	Protein NDRG4 OS=Homo sapiens OX=9606 GN=NDRG4 PE=1 SV=2
TRIM45	0	Tripartite motif-containing protein 45 OS=Bos taurus OX=9913 GN=TRIM45 PE=2 SV=1
MZB1	4.03E-95	Marginal zone B- and B1-cell-specific protein OS=Homo sapiens OX=9606 GN=MZB1 PE=1 SV=1
CASP14	5.17E-131	Caspase-14 OS=Homo sapiens OX=9606 GN=CASP14 PE=1 SV=2
UFM1	2.20E-44	Ubiquitin-fold modifier 1 OS=Pongo abelii OX=9601 GN=UFM1 PE=3 SV=1
DAW1	0	Dynein assembly factor with WDR repeat domains 1 OS=Bos taurus OX=9913 GN=DAW1 PE=2 SV=1
Tspear	0	Thrombospondin-type laminin G domain and EAR repeat-containing protein OS=Mus musculus OX=10090 GN=Tspear PE=1 SV=1
GUF1	0	Translation factor GUF1, mitochondrial OS=Bos taurus OX=9913 GN=GUF1 PE=2 SV=1
EGFL6	0	Epidermal growth factor-like protein 6 OS=Homo sapiens OX=9606 GN=EGFL6 PE=2 SV=1
GLT8D2	0	Glycosyltransferase 8 domain-containing protein 2

		OS=Bos taurus OX=9913 GN=GLT8D2 PE=2 SV=1
CFAP52	0	Cilia- and flagella-associated protein 52 OS=Homo sapiens OX=9606 GN=CFAP52 PE=1 SV=3
GABPB2	0	GA-binding protein subunit beta-2 OS=Bos taurus OX=9913 GN=GABPB2 PE=2 SV=2
CYB5D1	6.77E-96	Cytochrome b5 domain-containing protein 1 OS=Homo sapiens OX=9606 GN=CYB5D1 PE=2 SV=1
CIB4	1.10E-86	Calcium and integrin-binding family member 4 OS=Ovis aries OX=9940 GN=CIB4 PE=2 SV=1
Pkmyt1	0	Membrane-associated tyrosine- and threonine-specific cdc2-inhibitory kinase OS=Mus musculus OX=10090 GN=Pkmyt1 PE=2 SV=3
KPTN	0	KICSTOR complex protein kaptin OS=Homo sapiens OX=9606 GN=KPTN PE=1 SV=2
SLC5A6	0	Sodium-dependent multivitamin transporter OS=Oryctolagus cuniculus OX=9986 GN=SLC5A6 PE=2 SV=1
CHP1	7.22E-42	Calcineurin B homologous protein 1 OS=Pongo abelii OX=9601 GN=CHP1 PE=2 SV=3
RBM44	0	RNA-binding protein 44 OS=Bos taurus OX=9913 GN=RBM44 PE=3 SV=1
Cluap1	0	Clusterin-associated protein 1 OS=Rattus norvegicus OX=10116 GN=Cluap1 PE=1 SV=1
TMEM108	0	Transmembrane protein 108 OS=Bos taurus OX=9913 GN=TMEM108 PE=2 SV=1
unanme	4.29E-79	Uncharacterized protein C2orf50 homolog OS=Bos taurus OX=9913 PE=2 SV=2
ZNF207	0	BUB3-interacting and GLEBS motif-containing protein ZNF207 OS=Pongo abelii OX=9601 GN=ZNF207 PE=2 SV=1
ODAPH	1.25E-39	Odontogenesis associated phosphoprotein OS=Homo sapiens OX=9606 GN=ODAPH PE=1 SV=1
TOR3A	0	Torsin-3A OS=Homo sapiens OX=9606 GN=TOR3A PE=1 SV=1
PNKD	0	Probable hydrolase PNKD OS=Bos taurus OX=9913 GN=PNKD PE=2 SV=1
Rsph14	2.59E-54	Radial spoke head 14 homolog OS=Mus musculus OX=10090 GN=Rsph14 PE=2 SV=1
FAM227A	0	Protein FAM227A OS=Homo sapiens OX=9606 GN=FAM227A PE=1 SV=1
CTNS	0	Cystinosin OS=Bos taurus OX=9913 GN=CTNS PE=2 SV=1
BUD23	4.62E-63	Probable 18S rRNA (guanine-N(7))-methyltransferase OS=Bos taurus OX=9913 GN=BUD23 PE=2 SV=1

Tcf7	0	Transcription factor 7 OS=Mus musculus OX=10090 GN=Tcf7 PE=2 SV=2
FAM189A1	0	Protein FAM189A1 OS=Homo sapiens OX=9606 GN=FAM189A1 PE=2 SV=4
ATXN7L1	0	Ataxin-7-like protein 1 OS=Homo sapiens OX=9606 GN=ATXN7L1 PE=1 SV=3
HTATIP2	3.10E-153	Oxidoreductase HTATIP2 OS=Homo sapiens OX=9606 GN=HTATIP2 PE=1 SV=2
IQCD	0	Dynein regulatory complex protein 10 OS=Bos taurus OX=9913 GN=IQCD PE=2 SV=1
SPATA6	0	Spermatogenesis-associated protein 6 OS=Bos taurus OX=9913 GN=SPATA6 PE=2 SV=2
CHRNA9	0	Neuronal acetylcholine receptor subunit alpha-9 OS=Homo sapiens OX=9606 GN=CHRNA9 PE=1 SV=2
CABP5	4.67E-102	Calcium-binding protein 5 OS=Homo sapiens OX=9606 GN=CABP5 PE=1 SV=1
SERTAD4	3.37E-128	SERTA domain-containing protein 4 OS=Homo sapiens OX=9606 GN=SERTAD4 PE=2 SV=1
TBC1D16	0	TBC1 domain family member 16 OS=Homo sapiens OX=9606 GN=TBC1D16 PE=2 SV=1
Klk11	5.18E-157	Kallikrein-11 OS=Mus musculus OX=10090 GN=Klk11 PE=2 SV=1
Kncn	3.67E-53	Kinocilin OS=Mus musculus OX=10090 GN=Kncn PE=1 SV=1
FAM131C	1.33E-131	Protein FAM131C OS=Homo sapiens OX=9606 GN=FAM131C PE=1 SV=2
TTLL1	0	Probable tubulin polyglutamylase TTLL1 OS=Homo sapiens OX=9606 GN=TTLL1 PE=2 SV=1
TGS1	0	Trimethylguanosine synthase OS=Homo sapiens OX=9606 GN=TGS1 PE=1 SV=3
MRPS15	1.28E-92	28S ribosomal protein S15, mitochondrial OS=Bos taurus OX=9913 GN=MRPS15 PE=1 SV=3
CCDC62	0	Coiled-coil domain-containing protein 62 OS=Homo sapiens OX=9606 GN=CCDC62 PE=1 SV=2
IQCF2	4.17E-59	IQ domain-containing protein F2 OS=Bos taurus OX=9913 GN=IQCF2 PE=2 SV=1
SLX4IP	2.34E-111	Protein SLX4IP OS=Homo sapiens OX=9606 GN=SLX4IP PE=1 SV=1
NOB1	0	RNA-binding protein NOB1 OS=Bos taurus OX=9913 GN=NOB1 PE=2 SV=1
AK6	1.73E-83	Adenylate kinase isoenzyme 6 OS=Bos taurus OX=9913 GN=AK6 PE=2 SV=1
PRDM10	0	PR domain zinc finger protein 10 OS=Homo sapiens OX=9606 GN=PRDM10 PE=1 SV=3

UNC80	0	Protein unc-80 homolog OS=Homo sapiens OX=9606 GN=UNC80 PE=1 SV=2
BEND4	0	BEN domain-containing protein 4 OS=Homo sapiens OX=9606 GN=BEND4 PE=2 SV=3
SLC24A1	0	Sodium/potassium/calcium exchanger 1 OS=Homo sapiens OX=9606 GN=SLC24A1 PE=1 SV=1
Pax9	0	Paired box protein Pax-9 OS=Mus musculus OX=10090 GN=Pax9 PE=2 SV=1
GHRHR	0	Growth hormone-releasing hormone receptor OS=Homo sapiens OX=9606 GN=GHRHR PE=1 SV=2
PPP4R4	0	Serine/threonine-protein phosphatase 4 regulatory subunit 4 OS=Homo sapiens OX=9606 GN=PPP4R4 PE=1 SV=1
ZMAT4	2.56E-126	Zinc finger matrin-type protein 4 OS=Bos taurus OX=9913 GN=ZMAT4 PE=2 SV=1
LYPLAL1	9.07E-136	Lysophospholipase-like protein 1 OS=Homo sapiens OX=9606 GN=LYPLAL1 PE=1 SV=3
DCAF17	0	DDB1- and CUL4-associated factor 17 OS=Homo sapiens OX=9606 GN=DCAF17 PE=1 SV=1
SIRT4	1.17E-169	NAD-dependent protein lipoamidase sirtuin-4, mitochondrial OS=Bos taurus OX=9913 GN=SIRT4 PE=2 SV=1
MMAA	0	Methylmalonic aciduria type A protein, mitochondrial OS=Homo sapiens OX=9606 GN=MMAA PE=1 SV=1
PIKFYVE	0	1-phosphatidylinositol 3-phosphate 5-kinase OS=Homo sapiens OX=9606 GN=PIKFYVE PE=1 SV=3
KIAA1549L	0	UPF0606 protein KIAA1549L OS=Homo sapiens OX=9606 GN=KIAA1549L PE=2 SV=2
Spock1	5.82E-170	Testican-1 OS=Mus musculus OX=10090 GN=Spock1 PE=2 SV=2
GTPBP1	0	GTP-binding protein 1 OS=Homo sapiens OX=9606 GN=GTPBP1 PE=1 SV=3
TMEM8B	0	Transmembrane protein 8B OS=Bos taurus OX=9913 GN=TMEM8B PE=2 SV=1
CA6	0	Carbonic anhydrase 6 OS=Canis lupus familiaris OX=9615 GN=CA6 PE=2 SV=1
SNX31	0	Sorting nexin-31 OS=Homo sapiens OX=9606 GN=SNX31 PE=1 SV=3
MGME1	0	Mitochondrial genome maintenance exonuclease 1 OS=Homo sapiens OX=9606 GN=MGME1 PE=1 SV=1
GPANK1	0	G patch domain and ankyrin repeat-containing protein 1 OS=Homo sapiens OX=9606 GN=GPANK1 PE=1 SV=1
IQCK	2.50E-141	IQ domain-containing protein K OS=Homo sapiens OX=9606 GN=IQCK PE=2 SV=1
FRMPD1	0	FERM and PDZ domain-containing protein 1 OS=Homo

		sapiens OX=9606 GN=FRMPD1 PE=1 SV=1
RPIA	0	Ribose-5-phosphate isomerase OS=Homo sapiens OX=9606 GN=RPIA PE=1 SV=3
CCDC38	0	Coiled-coil domain-containing protein 38 OS=Homo sapiens OX=9606 GN=CCDC38 PE=2 SV=1
NUTM1	0	NUT family member 1 OS=Homo sapiens OX=9606 GN=NUTM1 PE=1 SV=2
ERLIN2	0	Erlin-2 OS=Pongo abelii OX=9601 GN=ERLIN2 PE=2 SV=1
CD72	9.59E-135	B-cell differentiation antigen CD72 OS=Homo sapiens OX=9606 GN=CD72 PE=2 SV=1
BIRC6	0	Baculoviral IAP repeat-containing protein 6 OS=Homo sapiens OX=9606 GN=BIRC6 PE=1 SV=2
NR1H4	0	Bile acid receptor OS=Bos taurus OX=9913 GN=NR1H4 PE=2 SV=1
RELB	0	Transcription factor RelB OS=Homo sapiens OX=9606 GN=RELB PE=1 SV=2
TREML2	6.05E-126	Trem-like transcript 2 protein OS=Homo sapiens OX=9606 GN=TREML2 PE=1 SV=2
DYDC1	8.06E-75	DPY30 domain-containing protein 1 OS=Homo sapiens OX=9606 GN=DYDC1 PE=1 SV=1
Ninj2	2.46E-62	Ninjurin-2 OS=Rattus norvegicus OX=10116 GN=Ninj2 PE=2 SV=1
CLHC1	0	Clathrin heavy chain linker domain-containing protein 1 OS=Macaca fascicularis OX=9541 GN=CLHC1 PE=2 SV=1
RAB9B	1.33E-119	Ras-related protein Rab-9B OS=Pongo abelii OX=9601 GN=RAB9B PE=2 SV=1
SERPINE 3	1.33E-104	Serpin E3 OS=Homo sapiens OX=9606 GN=SERPINE3 PE=2 SV=2
FANCG	0	Fanconi anemia group G protein OS=Homo sapiens OX=9606 GN=FANCG PE=1 SV=1
SLC51B	4.93E-37	Organic solute transporter subunit beta OS=Homo sapiens OX=9606 GN=SLC51B PE=2 SV=2
VEGFA	1.29E-74	Vascular endothelial growth factor A OS=Canis lupus familiaris OX=9615 GN=VEGFA PE=1 SV=1
NKIRAS1	2.31E-66	NF-kappa-B inhibitor-interacting Ras-like protein 1 OS=Macaca fascicularis OX=9541 GN=NKIRAS1 PE=2 SV=1
PAX5	0	Paired box protein Pax-5 OS=Homo sapiens OX=9606 GN=PAX5 PE=1 SV=1
Gm11992	3.28E-144	Uncharacterized protein C7orf57 homolog OS=Mus musculus OX=10090 GN=Gm11992 PE=1 SV=1
HOXB1	3.15E-160	Homeobox protein Hox-B1 OS=Pan troglodytes OX=9598

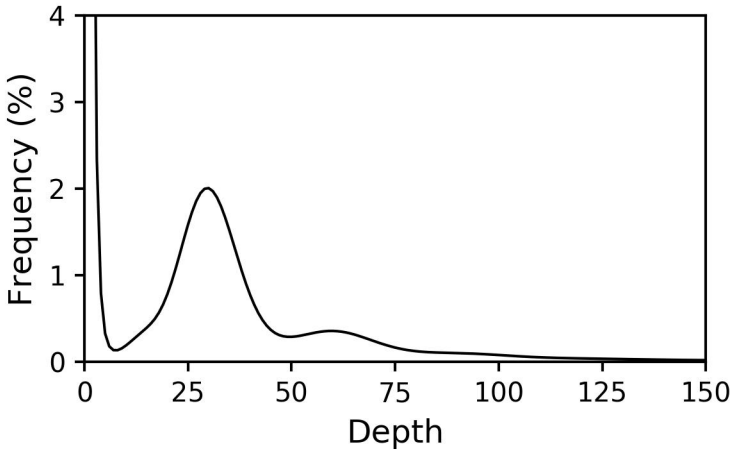
		GN=HOXB1 PE=3 SV=1
CIART	3.50E-149	Circadian-associated transcriptional repressor OS=Homo sapiens OX=9606 GN=CIART PE=1 SV=1
WDR33	0	pre-mRNA 3' end processing protein WDR33 OS=Homo sapiens OX=9606 GN=WDR33 PE=1 SV=2
RAB36	5.02E-169	Ras-related protein Rab-36 OS=Homo sapiens OX=9606 GN=RAB36 PE=2 SV=2
Tmprss3	0	Transmembrane protease serine 3 OS=Mus musculus OX=10090 GN=Tmprss3 PE=1 SV=2
E2F5	0	Transcription factor E2F5 OS=Homo sapiens OX=9606 GN=E2F5 PE=1 SV=1

Supplementary Table S16. 9 Positive selection genes may be related to adaptation of *V. ferrilata* to low oxygen and high UV radiation

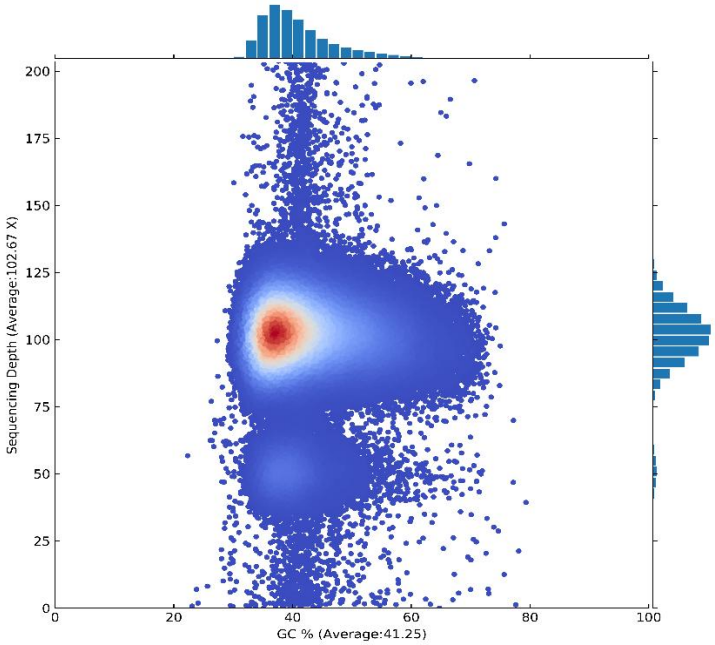
Function	Count	Genes
Angiogenesis related genes	6	Vascular endothelial growth factor A (*VEGFA*), RWD domain containing 3 (*RWDD3),*MYC associated factor X (*MAX*), HIV-1 tat interactive protein 2 (*HTATIP2*), EGF like domain multiple 6 (*EGFL6*), Placental growth factor (*PGF*)
DNA damage repair related genes	3	Mitochondrial genome maintenance exonuclease 1 (*MGME1*), FA complementation group G (*FANCG*), Sirtuin 4 (*SIRT4*)

Supplementary Table S17. Accession number.

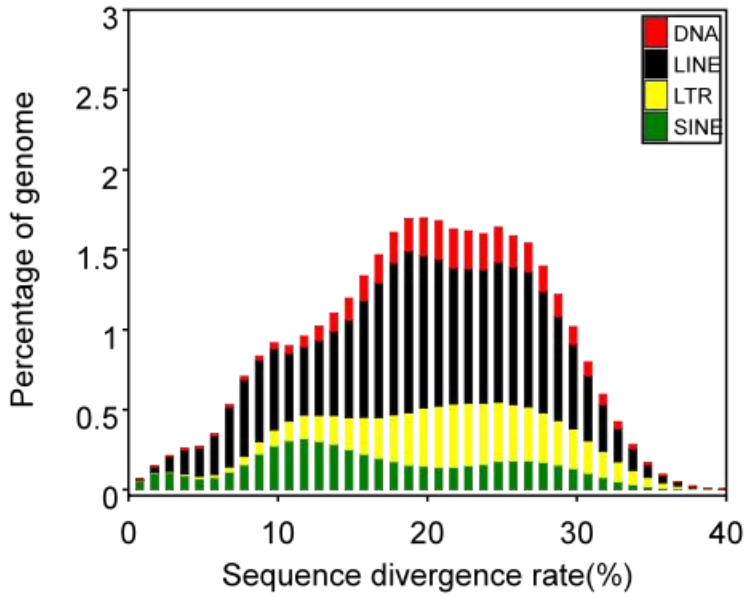
Category	App	Accession number
Genome and RNA sequencing data	NCBI Sequence Read Archive (SRA)	PRJNA762184
Hi-C sequencing data	NCBI Sequence Read Archive (SRA)	PRJNA768296
Genome assembly	NCBI WGS	JAJBZS000000000



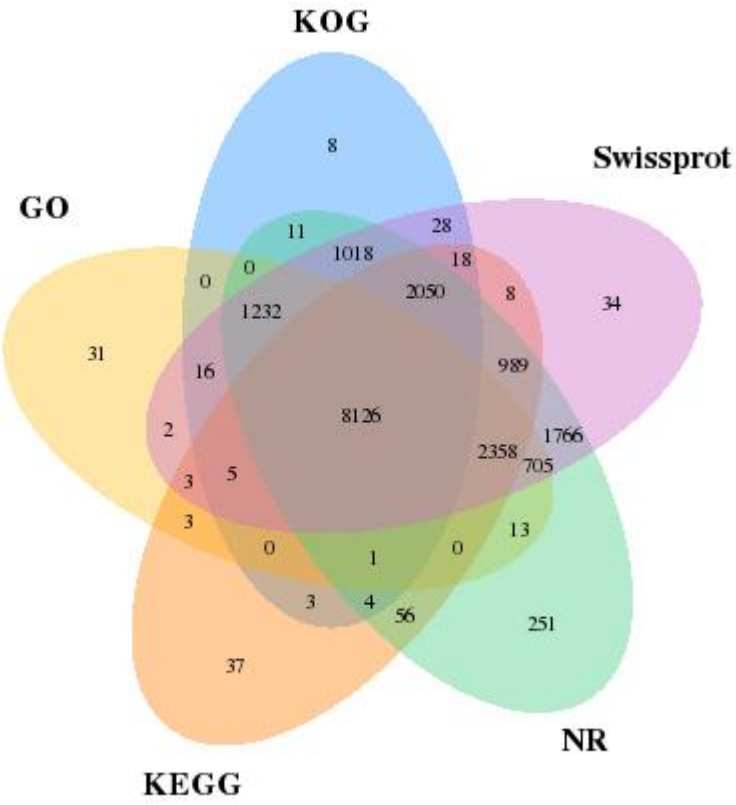
Supplementary Figure S1. Genome survey of *Vulpes ferrilata* using 17-mer analysis. K-mer depth on the horizontal axis and K-mer depth frequency on the vertical axis.



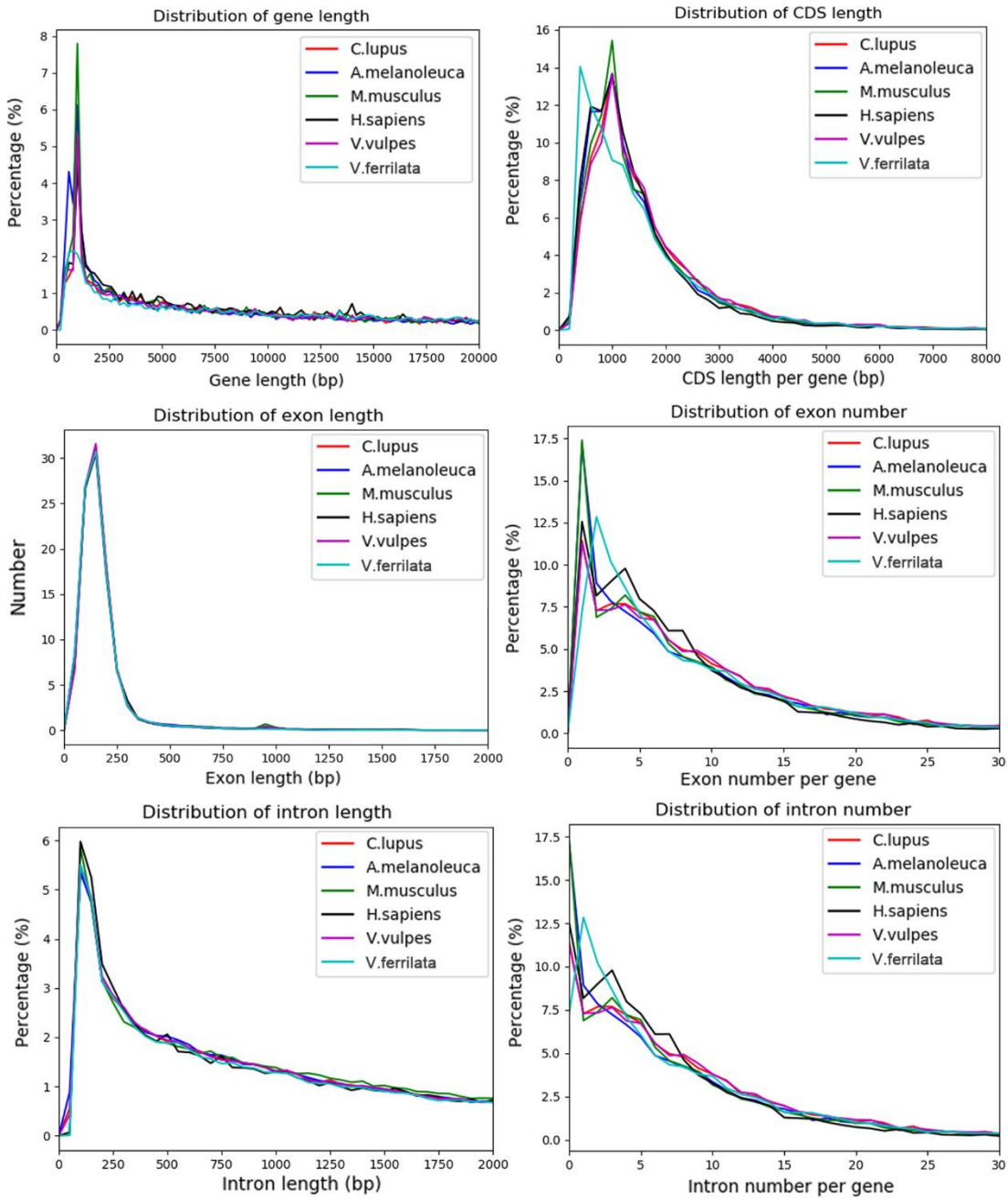
Supplementary Figure S2. The GC Depth distribution of the genome, where the abscissa is the GC content, and the ordinate is the Depth. These two values are counted sequentially in a 10Kb window.



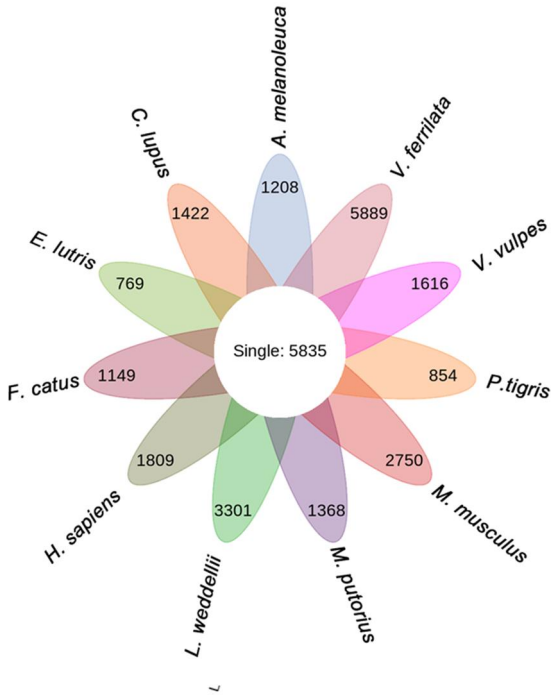
Supplementary Figure S3. RepeatableModeler transposon differentiation rate. The abscissor is the divergence degree between annotated TE sequences in the genome and corresponding sequences in the total library file, and the ordinate is the percentage of TE sequences under the divergence degree in the genome.Different TEs are labeled in different colors: DNA-red, LINE-black, LTR- yellow, and SINE-green.



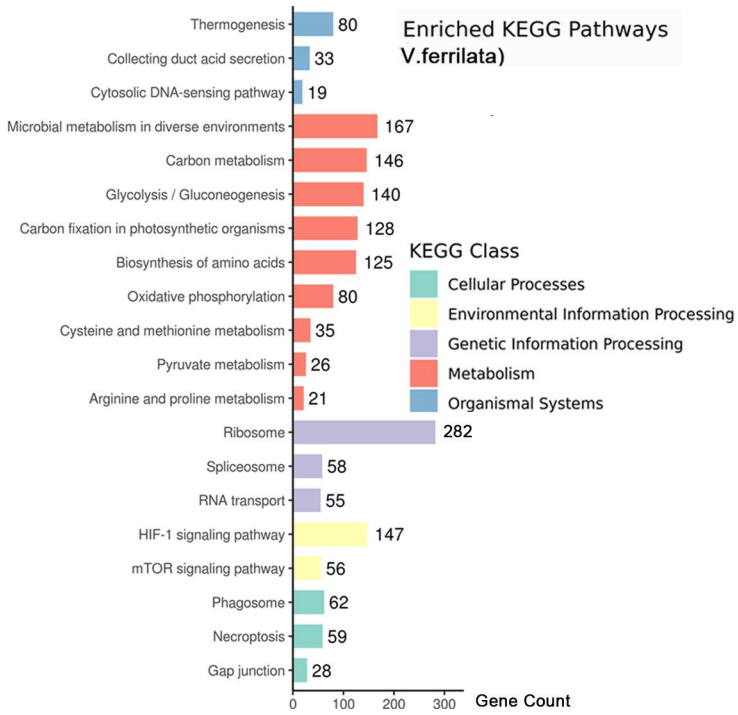
Supplementary Figure S4. The overlapping statistical mapping of Venn plot based on the gene function annotation results.



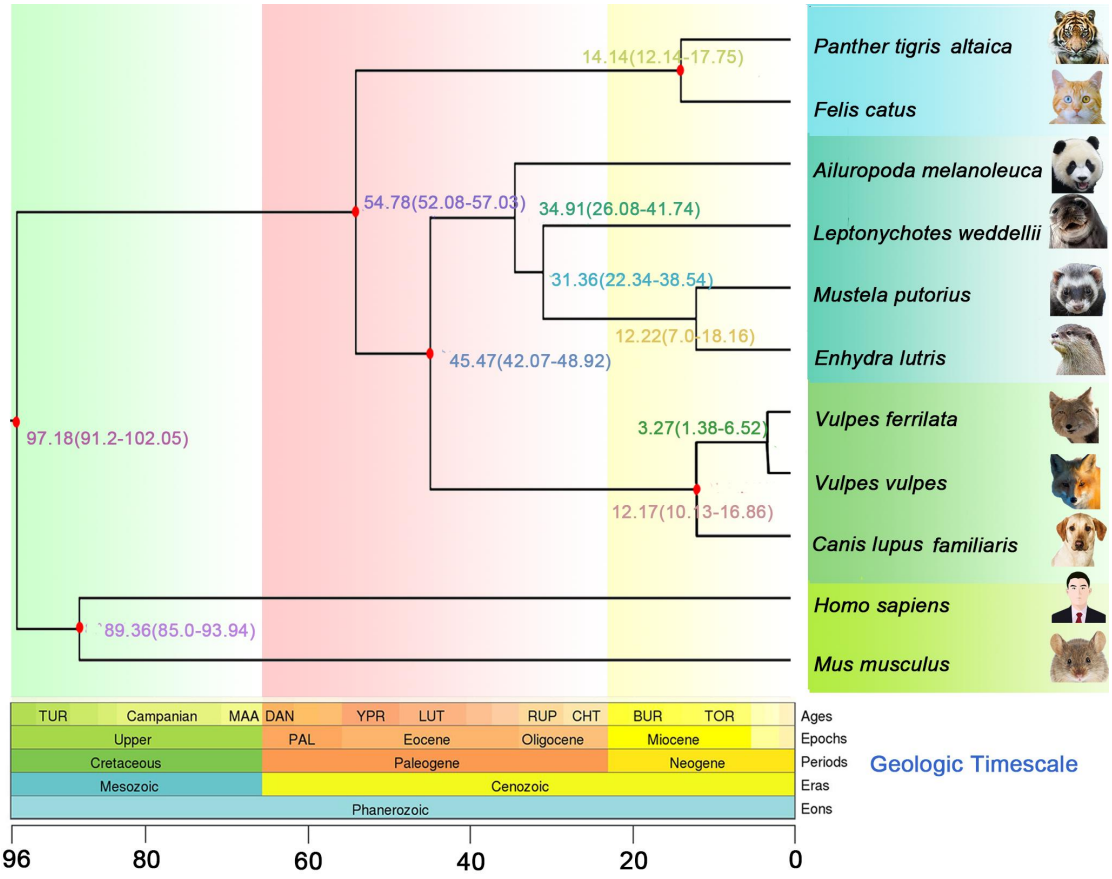
Supplementary Figure S5. Comparisons of the distribution of gene, coding sequences (CDS), exon and intron length for protein-coding genes in the genomes of Vulpes ferrilata and other species.



Supplementary Figure S6. Venn plot of single copy genes and specific genes of each species.



Supplementary Figure S7. KEGG enrichment analysis of species-specific genes of Vulpes ferrilata.



Supplementary Figure S8. Phylogenetic tree of *Vulpes ferrilata* with 10 genomes including 8 other Canidae. The value on each node represents the differentiation time of the present (Million years ago, Mya), the nodes marked in red are known fossil time points.