

Estimating Population Mutation and Error Rates from a Single Shotgun-Sequenced Diploid Bactrian Camel Genome

Pamela Burger^{1*}, Alessandra Palladini², Viola Nolte¹, Christian Schlötterer¹

¹Institute of Population Genetics, Vetmeduni Vienna, Veterinaerplatz 1, 1210 Vienna, Austria

²Università degli Studi di Pavia, D. to di Biologia Animale, Via A. Ferrata 1, 27100 Pavia, Italia

*Corresponding author: Pamela.Burger@vetmeduni.ac.at; T/F +43125077-4333/-4390

During domestication livestock animals have encountered extensive natural and artificial selection. Domestic animals, including camelids, therefore represent a unique resource to understand the genetic basis of phenotypic variation and genetic adaptation. The development of next-generation sequencing (NGS) technologies tremendously advanced whole-genome analyses making it possible to identify the changes underlying the process of domestication. Contrary to other livestock species for which the genome is already known (cattle, pig, chicken) or currently studied (sheep, alpaca, dromedary (Al-Swailem et al. 2010)) nuclear genomic data from the domestic Bactrian camel (*Camelus bactrianus*) have been missing so far. The aim of this study was to estimate the population mutation rate ($\theta = 4N_e\mu$) using the genome of a single Bactrian camel (Zoo Herberstein, Austria). We generated 20 Gigabases (Gb) of raw reads by paired-end read sequencing (2x101 bp) on a Illumina Genome Analyzer Iix. First, we made a *de novo* assembly and obtained nearly 2 Gb genomic sequence (min./ max. contig size = 100/ 29,951 bp; N50 = 1490 bp), which corresponds to almost two thirds of the Bactrian camel genome. In the next step, we mapped the reads with BWA 0.5.7 (Li & Durbin 2009) against the *de novo* assembled Bactrian camel genome. With an average 5.3-fold sequence coverage we discovered 304,232 polymorphic single nucleotide polymorphisms (SNPs) (SAMtools; Li et al. 2009). Using mlRho (Haubold et al. 2010) we obtained a likelihood estimation of the population mutation rate θ of 1.29×10^{-3} with a sequencing error rate of 6.64×10^{-4} . Compared to other domesticated ungulates our observed nucleotide diversity is higher than in cattle (9.4×10^{-4} ; Eck et al. 2009) but similar to pig ($1.1 - 2.1 \times 10^{-3}$; Amaral et al. 2009). Our results provide a start for the investigation of selection processes during camel domestication and for future association studies.

Keywords: SNPs, NGS, *camelidae*, effective population size

Al-Swailem AM, Shehata MM et al. (2010) Sequencing, analysis, and annotation of expressed sequence tags for *Camelus dromedarius*. *PLoS One*, 5, e10720.

Amaral AJ, Megens HJ et al. (2009) Application of massive parallel sequencing to whole genome SNP discovery in the porcine genome. *BMC Genomics*, 10:374.

Eck SH, Bent-Pagès A et al. (2009) Whole genome sequencing of a single *Bos taurus* animal for single nucleotide polymorphism discovery. *Genome Biology*, 10:R82.

Haubold B, Pfaffelhuber P, Lynch M (2010) mlRho – a program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes. *Molecular Ecology*, 19 (Suppl.1), 277-284.

Kofler R, Orozco-terWengel P et al. (2011) PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One*, 6, e15925.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754-60.

Li H, Handsaker B et al. (2009) The Sequence alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078-79.