

**The effect of the choice of reference genome in studying  
the origin of parthenogenetic rock lizards *Darevskia armeniaca* and *Darevskia dahli***

**Nika Nikolaishvili**

*Master's thesis submitted to the Faculty of Natural Science and Medicine of the Illia State University according to the requirements for the Degree of Master of Science Applied Genetics*

Program: Applied Genetics

**Supervisor: Oleksiy Yanchukov, Associate Professor.**

**Ilia State University  
Tbilisi, 202**

**Table of Content**

Abstract/Research issue..... 1  
Introduction/The review of the literature..... 2  
The object and methods of research..... 4  
Results..... 7  
Discussion..... 14  
Conclusion..... 16  
Bibliography..... 17  
Appendix..... 21

## **Abstract/Research issue**

The Caucasian rock lizards *Darevskia*, is a remarkably diverse genus, which includes several parthenogenetic species of hybrid origin. The utility of genome-wide ddRAD(Double digest restriction-site associated DNA) markers in this group allows to unravel the origin and the genetic structure of parthenogenetic populations in unprecedented detail. As a first step in the analysis, we mapped short-read sequences to one of the reference genomes available: either maternal or paternal parent, or a distant outgroup *Podarcis muralis*. We investigated the effect of the choice of the reference genome (i.e. reference bias) on the perceived relationship among two parthenogens, *D. armeniaca* and *D. dahli*, as well as their parental species *D. portschisnkii*, *D. mixta* and *D. valentini*. We used, Bowtie2, Samtools, Stacks 2, RADpainter and IQ-tree to produce genetic co-ancestry matrices and phylogenies from SNP data, elucidating the proximity of parthenogenetic species to their assumed parental species. Our analysis underscores the impact of reference genome choice on the perceived strength of species relationships. As the availability of genomes grows, so does the risk of biased results, underscoring the criticality of selecting the most suitable reference genome in genomic investigations, especially in cases that focus on interspecific hybrids, such as parthenogenetic species of *Darevskia* or *Aspidoscelis*

## **Introduction/The review of the literature**

DNA sequencing technology was first described by Sanger et al. in 1977, using chain terminating inhibitors(Sanger, Nicklen, and Coulson 1977). Although revolutionary for its own time, this method had limited throughput and high costs. Since then, new sequencing technologies have been developed and improved, not only increasing the quality and the length of the reads, but also significantly reducing its cost.(Kumar, Cowley, and Davis 2024, Hu et al. 2021). As a result, there has been an increasing number of research dedicated to