



INTRODUCING A FORENSIC VERSION OF PHYLOTREE FOR IMPROVED HAPLOGROUPING



Arne Dür¹, Nicole Huber² and Walther Parson^{2,3}

¹ Institute of Mathematics, University of Innsbruck, Innsbruck, Austria

² Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria

³ Forensic Science Program, The Pennsylvania State University, University Park, PA, USA

nicole.huber@i-med.ac.at



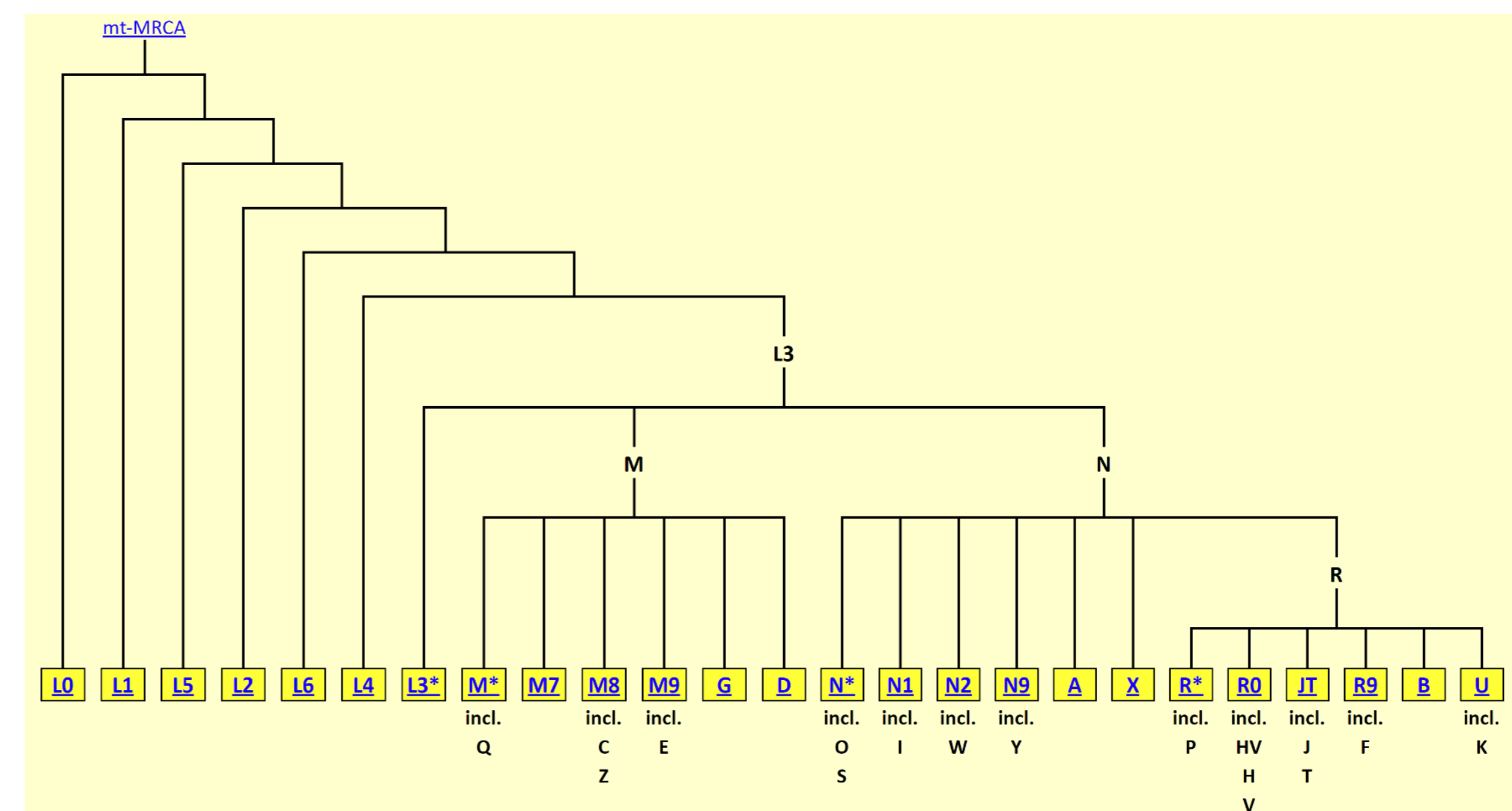
Co-funded by the Internal
Security Fund of the
European Union

dna.bases

Introduction

Mitochondrial DNA haplogrouping is an important tool that is not restricted to forensic mitochondrial genome (mitogenome) analysis. In many population genetic studies, haplogroup determination is a decisive prerequisite for further investigations and research questions.

In general, haplogroup determination is based on the nomenclature provided by PhyloTree [1]. The latest version (build 17) uses 24,275 full mitogenomes, resulting in a tree with 5435 different haplogroup defining motifs. However, an increasing number of sequences do not truly fit into the tree and numerous existing variants have not been used to characterize the respective clades.



Materials and Methods

Based on 24,928 reviewed full mitogenomes, we refined the original haplogroup motifs of Phylotree 17 without changing the original haplogroup nomenclature. Evaluation was conducted on an internal data set containing more than 100,000 mitogenomes as well as on previously published data.



Results

1. Modification of PhyloTree17 motifs:

About 750 motifs were modified through realignment by SAM2 [2] through haplogroup specific modifications and addition of frequently observed variants.

2. Addition of new motifs:

In total, 878 potential subclades were added. Each subclade was supported with at least two representing genomes.

In summary we have 6313 (5435+878) motifs for alignment and haplogrouping which is an increase of about 16% compared to Phylotree Build 17. For the control region that is commonly used in forensic science the increase from 2651 to 3443 different motifs is even 30%.

Discussion

The extension of Phylotree contributes to an improved resolution of the human mtDNA tree. Moreover, a more precise haplogroup assignment for both the full mitogenome and its control region was achieved.

References

- [1] van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30(2):E386-E394.
[2] Huber N, Parson W, Dür A. 2018. Next generation database search algorithm for forensic mitogenome analyses. *FSI Genet.* 2018 Nov;37:204-214.

Acknowledgements

This work received support from the European Union grant agreement number 779485-STEFA - ISFP-2016-AG-IBA-ENFSI. The computational results presented have been achieved in part using the HPC infrastructure LEO of the University of Innsbruck

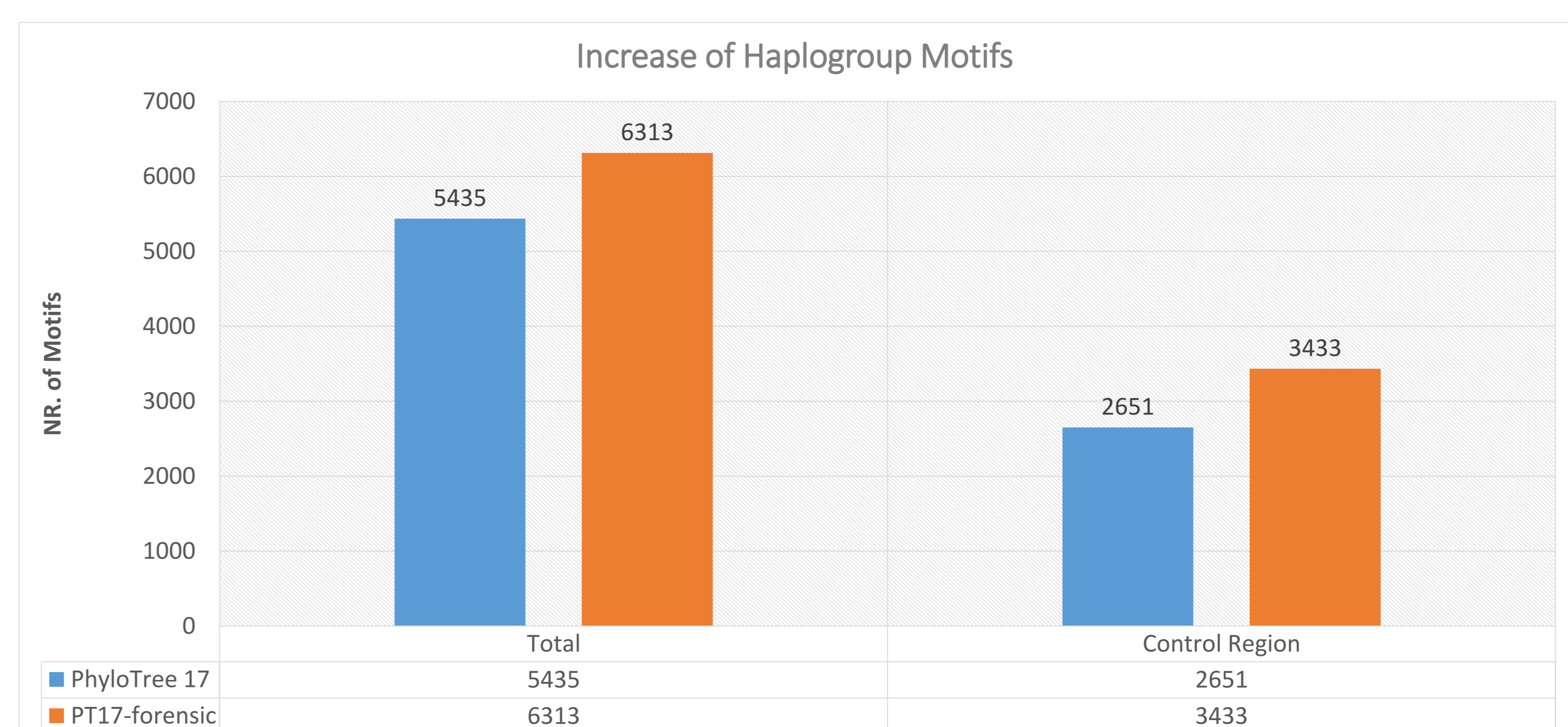


Figure 1: Increase of haplogroup motifs. Left pane: increase of full mitochondrial genomes. Right pane: increase of control region genomes.

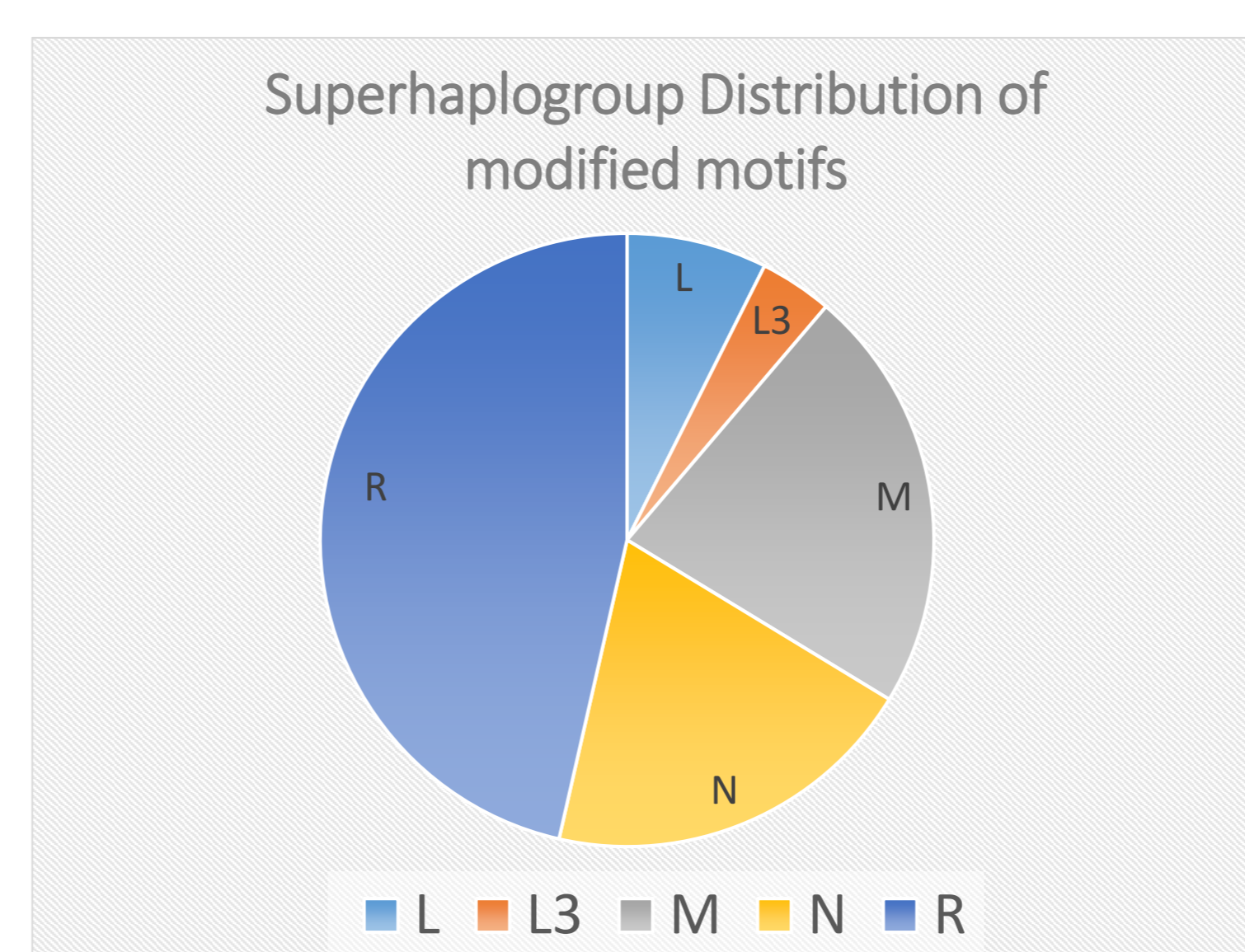


Figure 2: Distribution of superhaplogroups of modified motifs.

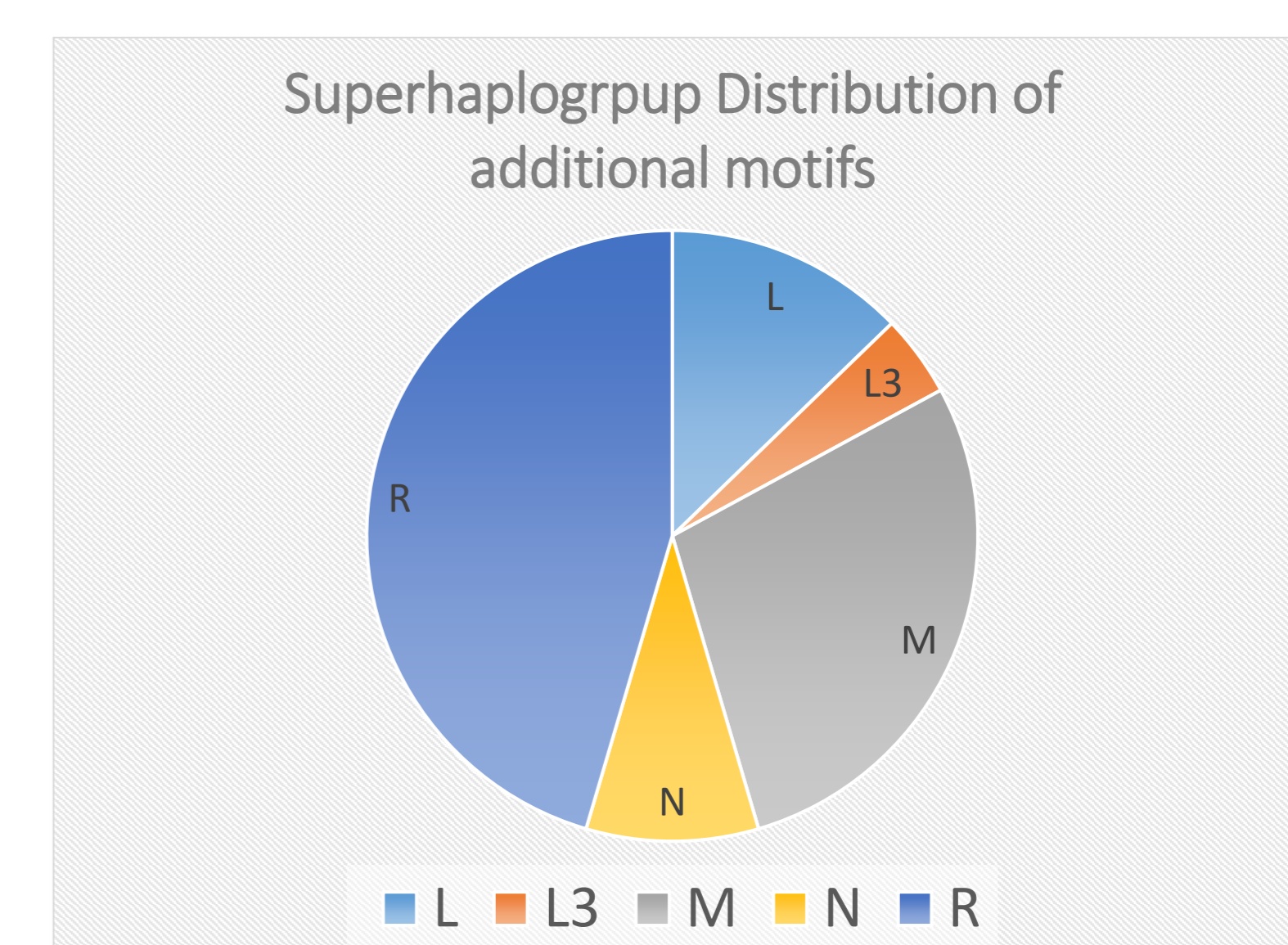


Figure 3: Distribution of superhaplogroups of newly added motifs.