Biotechnology is dependent upon the extraordinary efficiency, specificity, and versatility of enzyme function. Over the last decade, the revolution in sequencing technologies has produced vast amounts of sequence information from diverse biological sources. However, we have few functional details about the majority of this data, and therefore have only harnessed a minute fraction of the repertoire of enzymes and metabolic pathways available in Nature. Strategies to predict and characterize the functions of unexplored sequence space are urgently needed. Here, we present an innovative approach to characterize and classify sequence, structure, and functional diversity of a diverse group of enzymes - the FMN-dependent nitroreductase superfamily. This superfamily is comprised of biotechnologically important enzymes¹, yet only a small number of enzymes have been characterized. We undertook a comprehensive analysis, using a unique combination of sequence, structural, functional and phylogenetic characterizations (>24,000 sequences, 54 structures and >10 enzymatic functions) to create the first global view of the nitroreductase superfamily² – of particular interest for biomedical, bioremediation, and biocatalysis applications. The superfamily was delineated into 22 distinct subgroups, 8 of which have no currently known function. Furthermore, we identified three “hot spots” within the nitroreductase scaffold that form the structural basis for the evolution of function, and revealed the key functional residues that have led to evolutionary adaptation through active site profiling. This information is instrumental to the rational redesign of the nitroreductase scaffold. We applied our new knowledge of the nitroreductase superfamily to screen >7,000 metagenomes from public and private repositories to expose the true diversity of NTR enzymes, this approach resulted in an extensive final dataset of ~1M novel nitroreductases. Prominent and subgroup specific enrichment profiles for distinct metagenomic environments were also revealed by subgroup profiling. To further investigate this newly discovered sequence space, we are performing large scale enzymatic activity profiling (>400 enzymes) to provide functional data on a vast number of novel nitroreductase enzymes, and develop an innovative “nitroreductase toolbox”, with wide-ranging potential for biotechnological applications.

². Akiva, Copp et al., submitted.