GID1A Mutagenesis

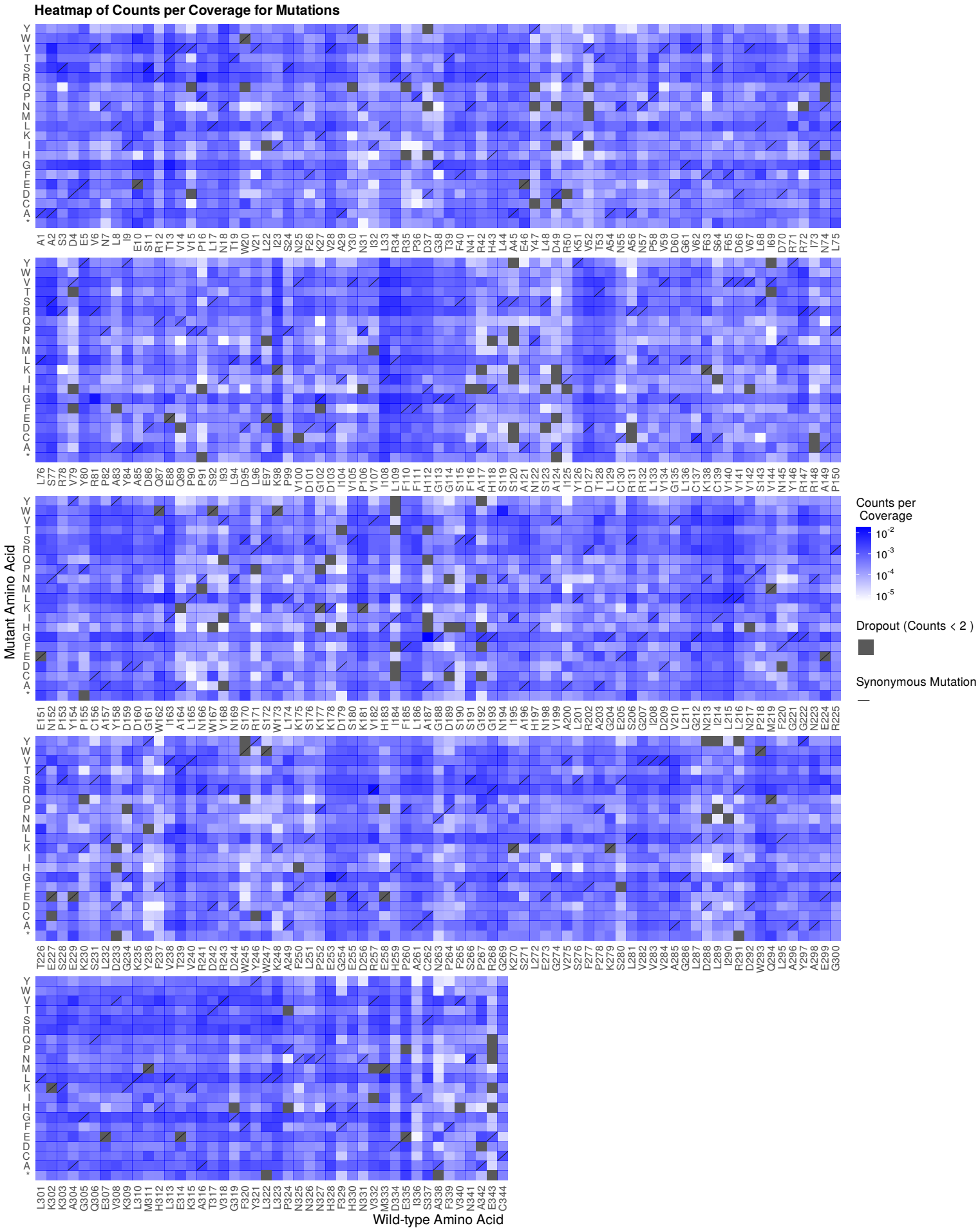
DMScore QC-Plots

Benjamin Wehnert[1], Taylor L. Mighell[1], Ben Lehner[1,2], Maximilian R. Stammnitz[1]
[1]Centre for Genomic Regulation (CRG), Barcelona, Spain; [2]Wellcome Sanger Institute, Cambridge, UK

GID1A is a gibberellin receptor in *Arabidopsis thaliana* that promotes plant growth by triggering DELLA protein degradation upon hormone binding. The 337-amino-acid protein was fully mutagenized using SUNi mutagenesis to generate 11,008 possible NNK/NNS single-codon variants. Variant frequencies were quantified via shotgun sequencing before and after growth competition. *DMScore* automatically processed the data, detecting over 95% of the designed variants. All QC plots shown below reflect DMScore's evaluation of the input library prior to selection.

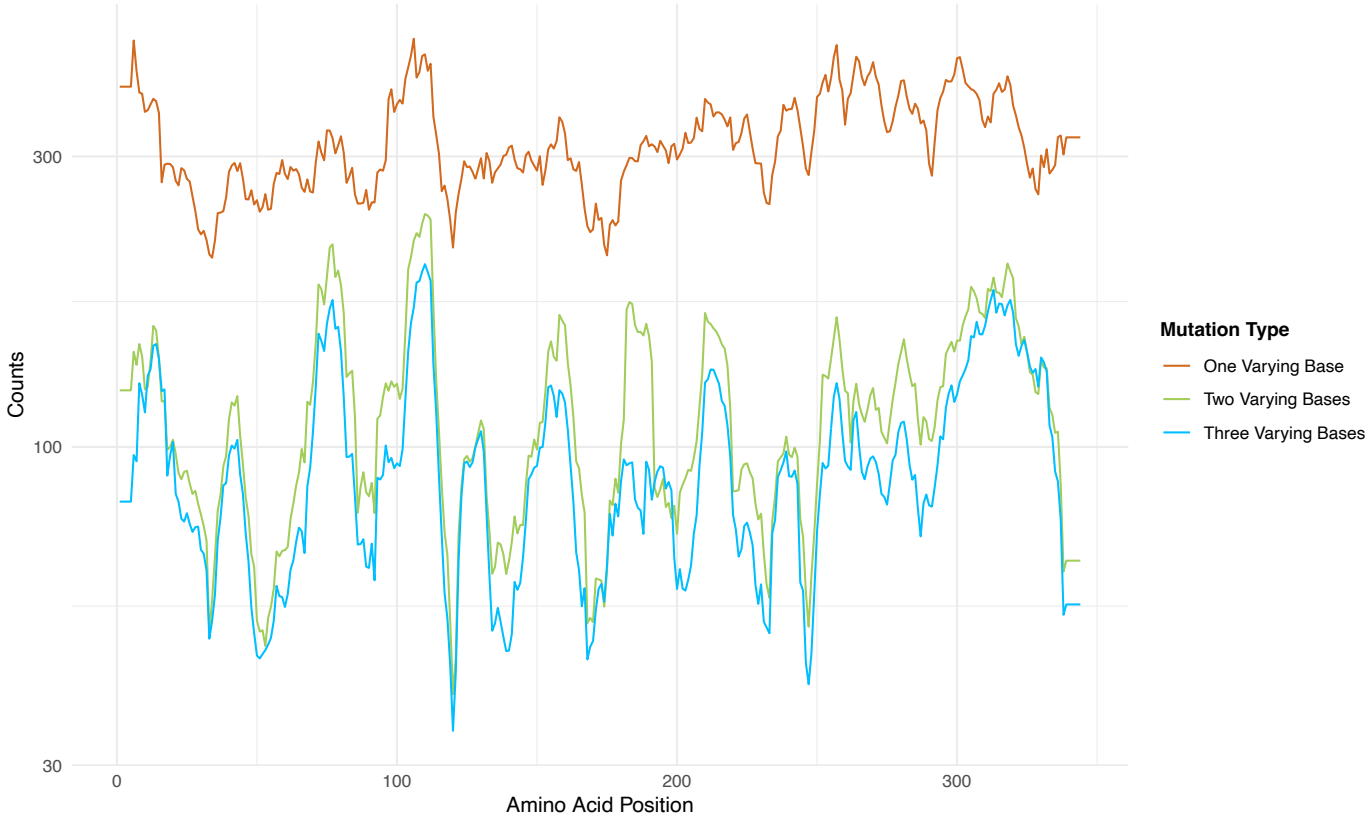
Variant Counts Heatmap

This heatmap visualizes the abundance of all single amino acid variants across the protein. Each cell represents one amino acid substitution at a specific position, with the color intensity showing the normalized count per sequencing coverage. Synonymous mutations are marked with diagonal lines. This plot holds detailed information to help validate the mutagenesis library.


Heatmap of Counts per Coverage for Mutations
Mutant Amino Acid
Wild-type Amino Acid
Counts per Coverage
10^{-2}
10^{-3}
10^{-4}
10^{-5}
Dropout (Counts < 2)
Synonymous Mutation

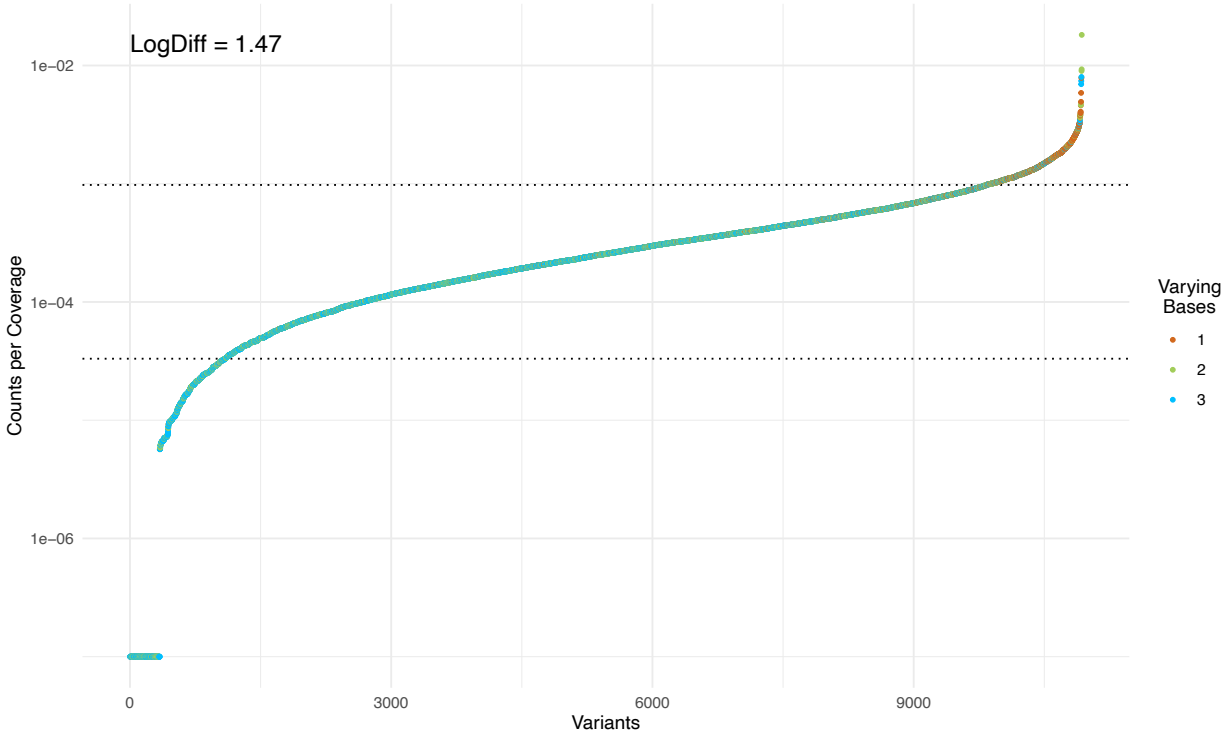
Position Biases

This plot summarizes the total variant counts per amino acid position, regardless of the specific mutation. It highlights intra-genic biases in coverage and mutational detectability across the protein sequence.



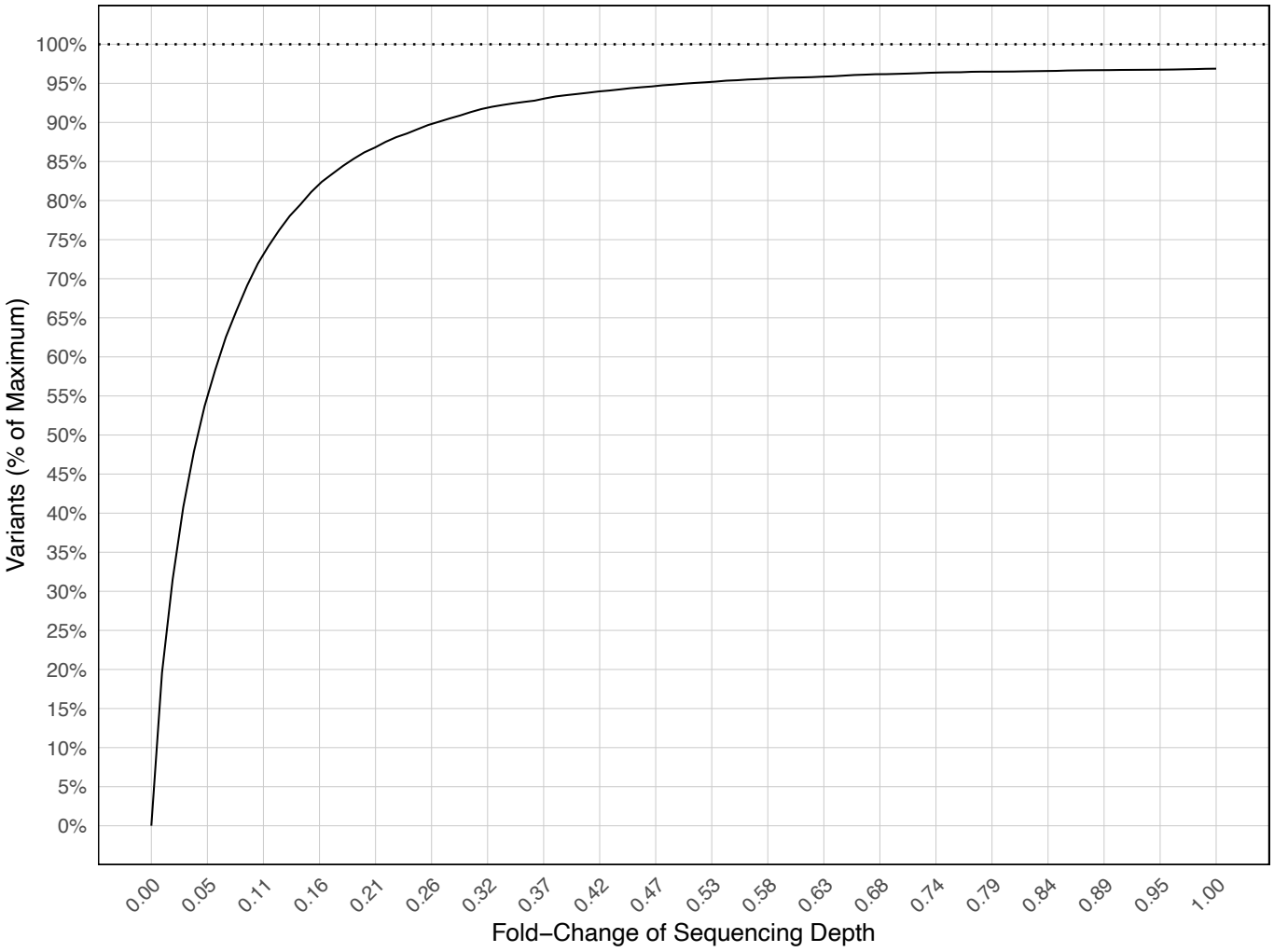
Variant Frequency Distribution

This plot ranks all variants by their abundance on a log scale. A steep slope indicates uneven variant representation, which may indicate biases in mutagenesis. The LogDiff value (90th – 10th percentile) summarizes this variation; a low LogDiff suggests a well-balanced library. The plot also shows biases toward the number of mutagenized bases. For nicking-based mutagenesis protocols, as used for this library, it is common to see a higher prevalence of single-base exchanges due to their chemically more efficient annealing.



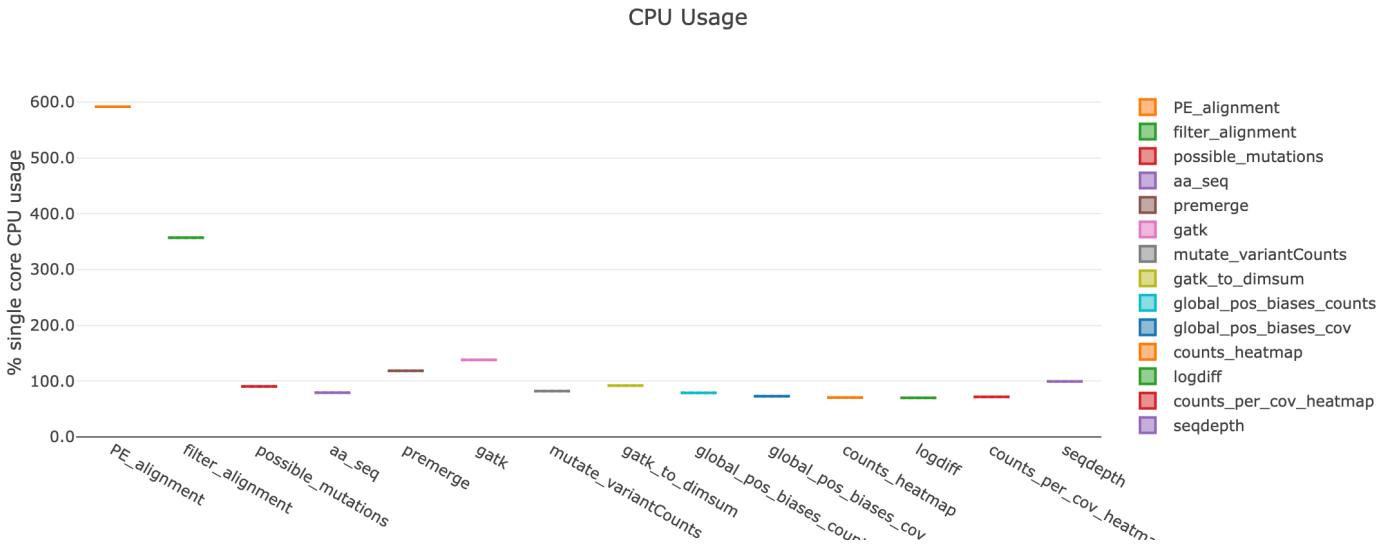
Sequencing Depth Evaluation

This plot assesses whether sequencing depth was sufficient to capture the variant diversity. It simulates how many variants would have been detected at different sequencing depths. A plateau near 100% indicates saturation; a rising curve suggests undersequencing.



Computing Ressources (by Nextflow)

This summary outlines the time and computational resources needed for each step in the DMScore pipeline. It helps assess workflow efficiency and informs future scalability.



Memory

Physical (RAM) | Virtual (RAM + Disk swap) | % RAM Allocated