FACULTAT
DE CIÈNCIES, TECNOLOGIA
I ENGINYERIES
UVIC | UVIC·UCC

Treball de Fi de Grau de Experimental

Exploring the role of histidine kinases in bacterial adaptation to drought

ANNA Mª COMAS PUJOL

Bachelor's Degree in Biotechnology

Tutor UVic: Marc Llirós Dupre

Tutor CEAB: Josep Ramoneda

Vic, June 2025

Agraïments

En primer lloc, m'agradaria donar les gràcies més sinceres a Josep Ramoneda i a en Marc Llirós, per la confiança, per la impecable tutorització i orientació que he rebut, i per tots els aprenentatges professionals que m'han transmès. També per la seva ajuda al llarg del projecte davant qualsevol dubte que he tingut, i per a fer la meva estada més amena en el CEAB. I a totes les persones que formen part del grup, per la proximitat i la manera en què m'han acollit i inclòs. A més, vull agrair a la UVic-UCC i al seu professorat per tots els coneixements i pensament crític que he pogut adquirir durant els anys de carrera, que estic segura que em seran de gran ajuda al meu futur.

Gràcies a les amistats fetes al llarg del grau, sense elles el pas per la universitat no hauria sigut el mateix. Finalment, gràcies a la meva família, per fer-me costat i aconsellar-me al llarg de tots els meus estudis. Doncs sense ells no hauria aconseguit tot el que m'he proposat. Gràcies per ser-hi sempre.

Resum

Títol: Exploring the role of histidine kinases in bacterial adaptation to drought

Autora: Anna Mª Comas Pujol

Co-Tutores: Marc Llirós Dupre (UVic) i Josep Ramoneda (CEAB)

Data: Juny de 2025

Paraules clau: Bioinformàtica, Adaptació bacteriana, Ecologia microbiana

El canvi climàtic està provocant una aridificació creixent dels ecosistemes, fet que posa en risc les funcions dels microbiomes del sòl, essencials per al cicle de nutrients i la biodiversitat. Comprendre com les bacteris s'adapten a la sequera és clau per predir la resiliència dels ecosistemes. Aquest treball parteix de la hipòtesi que les bacteris resistents i sensibles a la sequera presenten diferències en el nombre i diversitat de quinases d'histidina (HKs), proteïnes fonamentals per a la detecció de canvis ambientals. Els objectius han estat desenvolupar un pipeline bioinformàtic per identificar i anotar HKs en 1601 genomes bacterians classificats segons la seva resposta a la sequera, comparar la seva abundància i diversitat entre grups funcionals i explorar patrons taxonòmics. Es van descarregar els genomes, es van alinear i filtrar les seqüències de HKs i es van fer anàlisis estadístiques i comparatives. Els resultats mostren una gran variabilitat en el nombre de HKs per genoma (de 5 a 300), amb una tendència a major abundància en bacteris resistents a la sequera. Alguns fílums, com Proteobacteria i Actinobacteriota, presenten més HKs i diversitat, mentre que altres tenen valors menors, indicant que tant el llinatge evolutiu com l'ecologia influeixen en la distribució d'aquests sensors. La correlació entre la mida del genoma i el nombre de HKs no és determinant, apuntant a altres factors adaptatius. En conclusió, la diversitat i abundància de HKs reflecteixen l'adaptació bacteriana a condicions ambientals adverses com la sequera.

Summary

Title: Exploring the role of histidine kinases in bacterial adaptation to drought

Author: Anna Mª Comas Pujol

Supervisor: Marc Llirós Dupre (UVic) and Josep Ramoneda (CEAB)

Date: June 2025

Keywords: Bioinformatics, Bacterial adaptation, Microbial ecology

Climate change is driving increased aridification of ecosystems, threatening the structure and function of soil microbiomes that are essential for nutrient cycling and biodiversity. Understanding how bacteria adapt to drought is crucial for predicting ecosystem resilience. This study hypothesized that drought-resistant and drought-sensitive bacteria differ in the number and diversity of histidine kinases (HKs), key proteins in environmental sensing. To test this, we developed a bioinformatics pipeline to identify and annotate HKs in 1601 bacterial genomes previously classified by drought response. Genomes were downloaded and processed, HK sequences were aligned and filtered for quality, and statistical analyses were performed to compare HK abundance and diversity across functional and taxonomic groups. Results revealed high variability in HK number per genome (ranging from 5 to 300), with drought-resistant bacteria generally exhibiting higher HK counts, suggesting enhanced adaptive capacity. Taxonomic analysis showed that groups such as Proteobacteria and Actinobacteriota possess greater HK abundance and diversity, while other phyla have significantly lower values, indicating that both evolutionary lineage and ecological pressures shape HK distribution. Although some correlation exists between genome size and HK number, it is not decisive, highlighting the influence of additional adaptive factors. In conclusion, the diversity and abundance of HKs reflect bacterial adaptation to adverse environmental conditions like drought, with significant differences observed between functional groups and across bacterial lineages.

Content table

Tables Index

Figures Index

1. Introduction

Climate change is leading to extensive aridification across many areas of the planet (Tariq et al., 2024). These drylands cover approximately 41% of the global surface, and the expansion of arid lands has a huge impact on the structural and functional aspects of ecosystems (Feng & Fu, 2013). Drought limits water availability, putting the soil organisms in danger and stressing them, particularly forcing microorganisms to regulate osmotic conditions. Many of these microorganisms are important for the ecosystems by regulating nutrients and keeping biodiversity and microbiomes (Schimel, 2018). Microbiomes are of high interest,because they drive most of the biological transformations and their role in the development of the soil (pools of carbon (C), nitrogen (N), and other nutrients) make possible the growth of plant communities, among other functions (Philippot et al., 2021; Schulz et al., 2013). Since microbial communities need to invest energy to resist drought, this has important impacts on their functioning, as has been seen for example in carbon cycling (Malik & Bouskill, 2022). This acknowledges the importance of understanding how bacteria adapt to aridity and the description of the different strategies bacteria can follow.

The principal part of the ability to sense environmental changes is thanks to the two-component signal transduction system (TCS), which consists of a sensor histidine kinase (HK) and a cognate response (Figure 1) (Foster et al., 2004). The HK is an integral membrane protein with a homodimer formation it forms a 4-helical bundle in the membrane, with two TM helices from each monomer; with an extracellular sensor domain and an intracellular catalytic kinase domain. It also has dimerization and histidine phosphotransfer (DHp), the catalytic core where it takes place the three catalytic reactions: Histidine phosphorylation, phosphotransfer to the response regulator, and, for bifunctional HKs, a phosphatase reaction (not a simple reversal of the phosphotransfer reaction) (Bhate et al., 2015).

The functional procedure goes by binding a ligand or sensing a change in the environment, our HK undergoes autophosphorylation on the conserved histidine residue (Figure 2). This signal is transferred to the phosphoryl group from the HK to a conserved aspartate residue. When reaches the DHp this undergoes a modification on the H-box, which contains the site of autophosphorylation. Many HK have a phosphatase activity which dephosphorylates the response regulator and opposes kinase function. This response is often a transcription factor, that results in a change of the expression of a specific regulation that mediates on the adaptation of bacteria to the environmental changes sensed

by the HK. Although the specific responses mediated by certain TCSs have been elucidated, the actual ligands or signals that trigger histidine kinase autophosphorylation are often unknown (Wolanin et al., 2002).

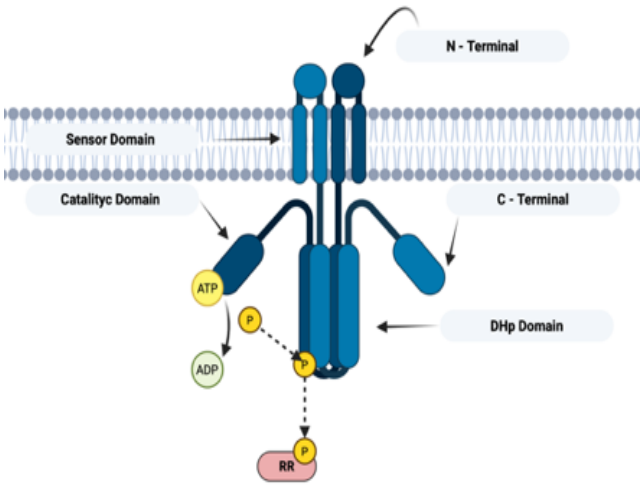


Figure 1 · Structure of a Histidine Kinase (Image created using BioRender)

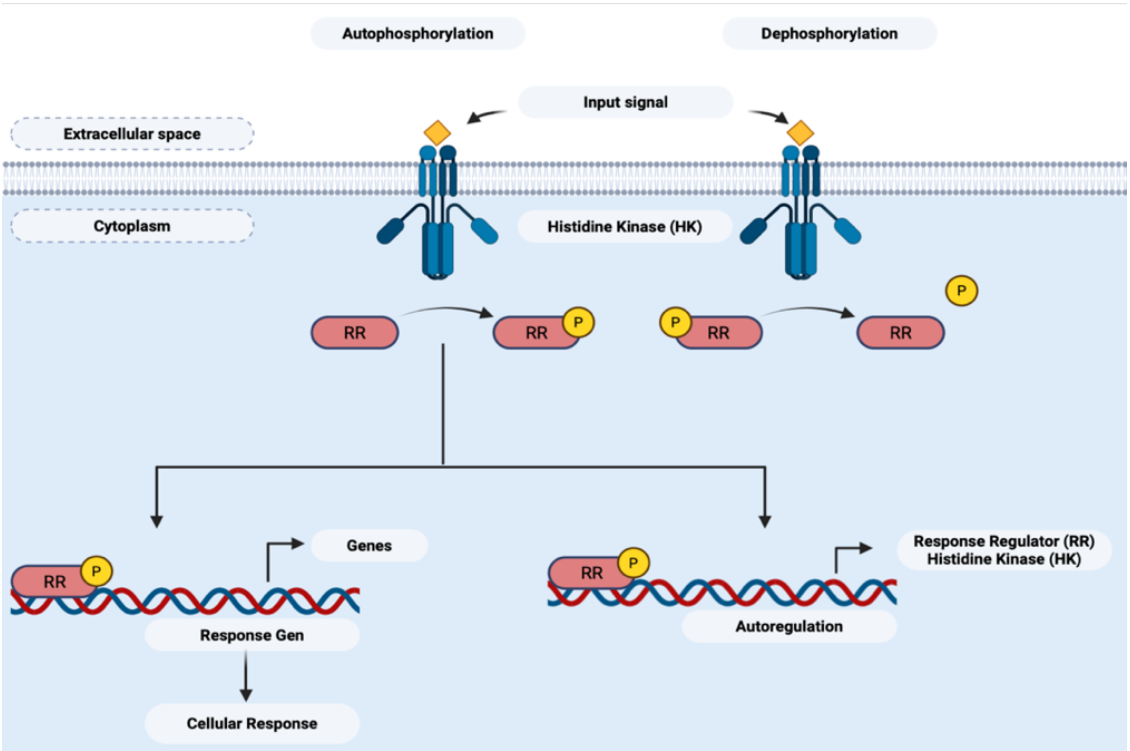


Figure 2 · Phosphotransferase Reactions (Image created using BioRender)

Several studies have demonstrated that HKs are central elements in the ability of bacteria to sense and respond to environmental changes (Louca et al., 2016). Drought, as an environmental stressor, involves alterations in multiple factors such as water availability, oxygen concentration, nutrients, and salts—all of which are signals that can be detected by HKs. Therefore, it is plausible to hypothesize that the composition and abundance of HKs in bacterial genomes are closely linked to their capacity to adapt to drought conditions (Tran et al., 2007). A recent study shows that the composition of HKs in microbiomes can differentiate radically different environments, providing evidence that HK diversity reflects microbial adaptation to specific environmental conditions. For this reason, analyzing the presence and variety of HKs in drought-resistant and drought-sensitive bacteria can offer valuable insights into the molecular mechanisms underlying microbial resilience in the face of increasing ecosystem aridification (Park et al., 2023).

2. Objectives

HKs are known to be involved in bacterial sensing of their surrounding environment. Therefore, we hypothesize that drought-resistant and drought-sensitive bacteria will have different types and numbers of HKs, reflecting their distinct adaptive strategies to environmental stress. Additionally, we propose that certain bacterial phyla will display higher HK abundance and diversity in their genomes than others, suggesting that evolutionary lineage and ecological niche influence the investment in signal transduction systems. We further hypothesize that the distribution of HKs in bacterial genomes is not solely determined by genome size, but is also shaped by ecological pressures and evolutionary history, leading to lineage- and environment-specific signaling repertoires. To test these hypotheses, we had the following objectives:

- Develop a pipeline to identify and annotate histidine kinases (HKs) of unknown function across a wide range of bacterial genomes.
- Compare the number and diversity of HKs present in genomes of drought-responsive versus drought-sensitive bacteria.
- Investigate whether different bacterial taxonomic groups (phyla) exhibit distinct patterns in the number and composition of HKs, which could reflect evolutionary and ecological adaptations.

3. Methodology

A first approach to the methodology in this work was represented schematically (Figure 3), and the information was explained in more detail in the following subsections.

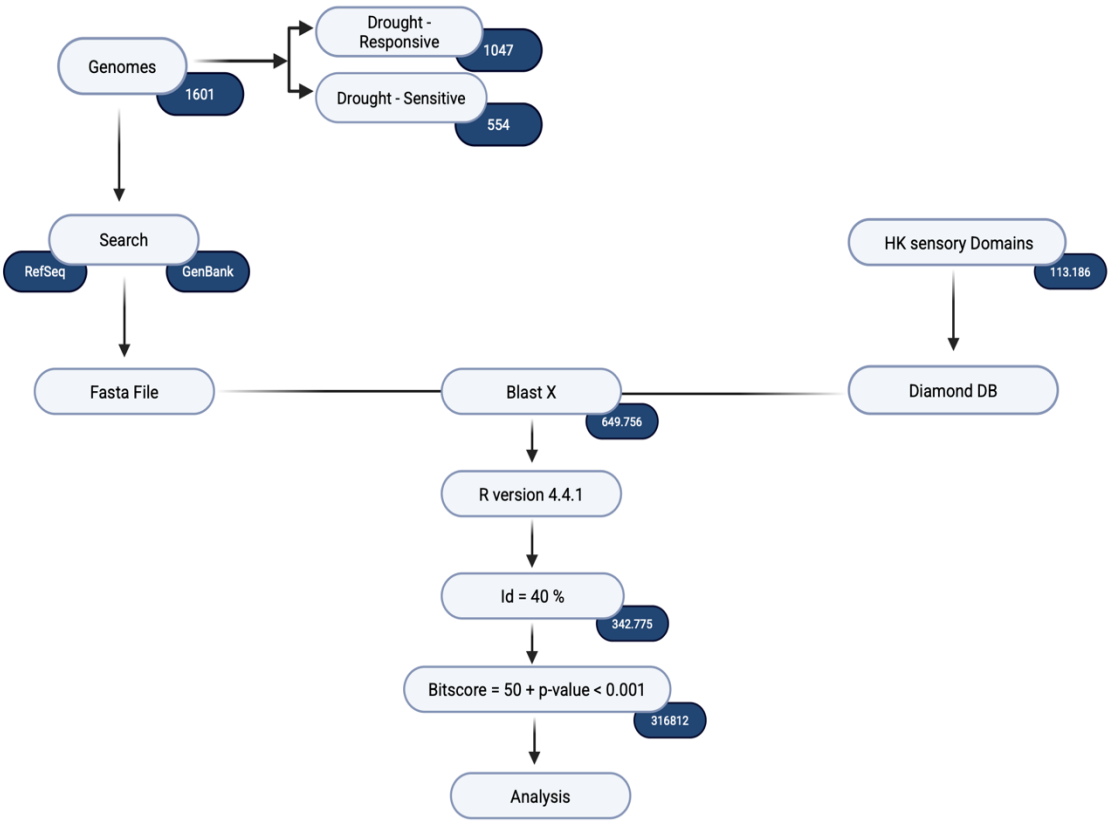


Figure 3 • Outline of the methodology, a global summary of the methodology and the gathering of the information. (Image created uding BioRendere)

3.1. Genome Data

We gathered 1,601 bacterial genomes, which had been previously classified as drought-responsive or sensitive based on their abundance within bacterial communities exposed to experimental drought. These genomes were classified according to their differential abundance in the drought treatment compared to the control, which received received normal precipitation in outdoor experimental plots. We identified 1,047 as drought-responsive and 554 as drought-sensitive. This information came from nine experiments, where half of the field was covered to simulate drought conditions, while the other half was left uncovered to receive normal precipitation. This experiment spanned a wide range of geographic places and climatic conditions. These bacterial communities had been characterized by sequencing 16S rRNA, already published, and a response classification by responsive/sensitive had been previously calculated on the group.

A preview of genome data showed that some phyla were predominant in the dataset, inlcuding Proteobacteria (35.5%), Actinobacteriota (27.4%), Firmicutes (4.1%) and Bacteroidota (10.1%). Additionally, five other phyla contained more than ten genomes: Acidobacteriota, Verrucomicrobiota, Myxococcota, Planctonycetota, and Chloroflexi; with the latter representing less than 1% of the genomes (Figure 4).

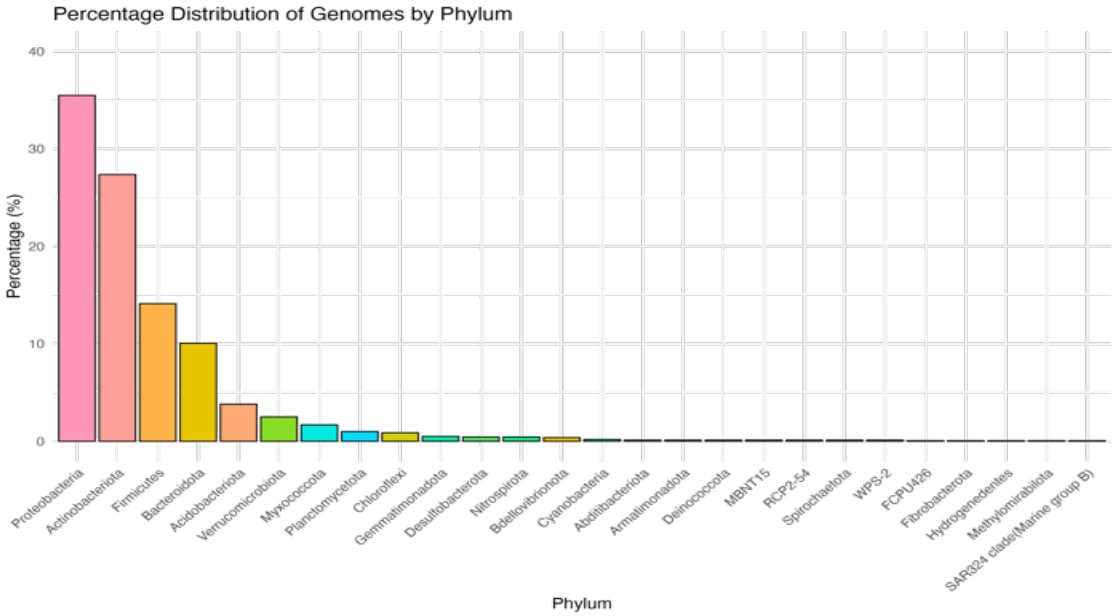


Figure 4 · Percentage distribution of genomes by phylum.

After adding a distribution by response groups, the three major representatives had more responsive than sensitive genomes, while the difference between the two groups remained small in the rest of the phyla (Figure 5). We observed a significant difference in the representation of the first four groups compared to the following ones, as the number of representatives declined.

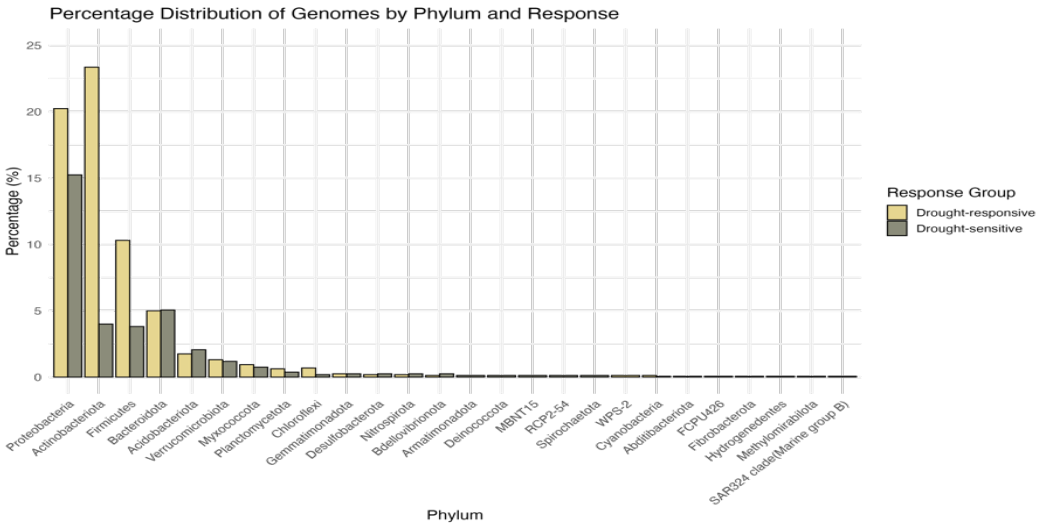


Figure 5 · Percentage distribution of genomes by phylum and response

3.2. Histidine kinase sequence information

From the article "A bacterial sensor taxonomy across earth ecosystems for machine learning applications" (Park et al., 2023), we took the amino acid sequences of the HKs, with the Pfam, gene name, and amino acid sequence for all clustered proteins. In this article, they identified HK proteins from 20,712 metagenomes from 76 different ecosystems almost 18,000,000 amino-acid sequences of unique sensors, which were clustered into domains with common function by Mmseqs2, ultimately resulting and in 113,186 HKs sensory domains.

We retrieved the sequence variants of the 113,186 sensory domains to create a new data base that could be used to annotate the bacterial genomes, we also added a new unique identifier (HK_xxxx) for each sequence, which was previously examined to verify all unique sequences matched the article's reported total of 113,186. We converted this list of HKs sensory domains into a Diamond Database, using the diamond-v2.1.10 package. This database was developed to reduce costs and computational time, as sequence alignment required significant processing power

3.3. Annotation of genomes into HKs

These 1,601 genomes were processed with R version 4.4.1, where we searched for these genome codes on the GenBank and RefSeq of the National Center for Biotechnology Information (NCBI), which belongs to the National Institute of Health (NIH) and created a document with the ID and the link to download the information. This document was sent to the cluster to download all the Fasta files, which were later unzipped. This yielded 1,601 fasta files, which were processed with a BlastX to align from protein to DNA and annotate using our Diamond database. We ran blast in sensitive mode with a minimum e-value of 0.001 and iterative alignment.

The results of this annotation producced one document for each genome, so we merged all the outputs in one file to do the analysis. The headers were added afterward using the Blastn output format 6 (Table 1). This header contained twelve columns (Metagenomics - BLASTn output format 6, n.d.).

Table 1 • Blastn output format 6. (Metagenomics - BLASTn output format 6, n.d.)

qseqid	query or source (gene) sequence id
sseqid	Subject or target (reference genome) sequence id
pident	Percentage of identical positions
lenght	Alignment length (sequence overlap)
mismatch	Number of mismatches
gapopen	Number of gap openings
qstart	Start of alignment in query
qend	End of alignment in query
sstart	Start of alignment in subject
send	End of alignment in subject
evalue	Expect value
bitscore	Bit score

We opened the output generated by Blastx with R version 4.4.1. We did filtered the data using specific criteria: sequence Identity (percentage identity) ≥ 40%, the identity value stipulated that two sequences with more than 30% identical over their length were homologous (Pearson, 2013); plus a > 40% identity indicated identical functions and conserved structural/active-site residues (Sangar et al., 2007). The E-value was < 0.001, which indicated the probability of finding good matches between our genomes and the numbers of sequences, a lower number indicated less casual march and ensured that it was a real match. The bit score (alignment quality) was ≥ 50; providing a double check and reaffirming that all the sequences were most likely homologous. After applying these three filters, we went from 649,756 HKs-genome matches to 316,812, excluding 51,24% of them. This final filtered data was merged with additional information from these genomes such as genome size, response to drought (responsive vs sensitive), or the taxonomic affiliation of the genomes. In this way, we created a database where bacterial responses to drought could be investigated for their association with HKs.

To visualize the distribution and relation of the HKs across the genomes, we generated several plots. Some of the plots are to see how many genomes contained each HKs also the unique HKs in each genome. Additionally, we conducted a correlation analysis between the HKs length and the number of genomes containing each HKs using Pearson's correlation test. All plots and statistical analyses downstream were performed using R (R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/).

4. Results and Discussion

4.1. General Distribution of HKs across genomes

A first approach to the data was made by visualizing the distribution of the HKs on the different genomes to examine the prevalence of HKs that appeared in several genomes. We saw a detailed representation in Annex A (Supplementary Figure (SP) 1). Notably, a substantial proportion of the HKs was found in fewer than 500 genomes, where rare HKs variants could be observed (Figure 6).

Our dataset contained 16,601 HKs; within this; we identified 6,697 unique HKs, where the majority were found in fewer than teen genomes. This pattern indicated that most HKs were either rare or functionally specialized, whereas those that appeared in numerous genomes represented essential or highly conserved biological roles. There was a functional census that showed that the majority of TCS proteins, including HKs, were present in only a few genomes, while a smaller core set was conserved across many taxa. The paper discussed the prevalence of rare, lineage-specific HKs versus a handful of universally conserved ones (M. Galperin, 2005).

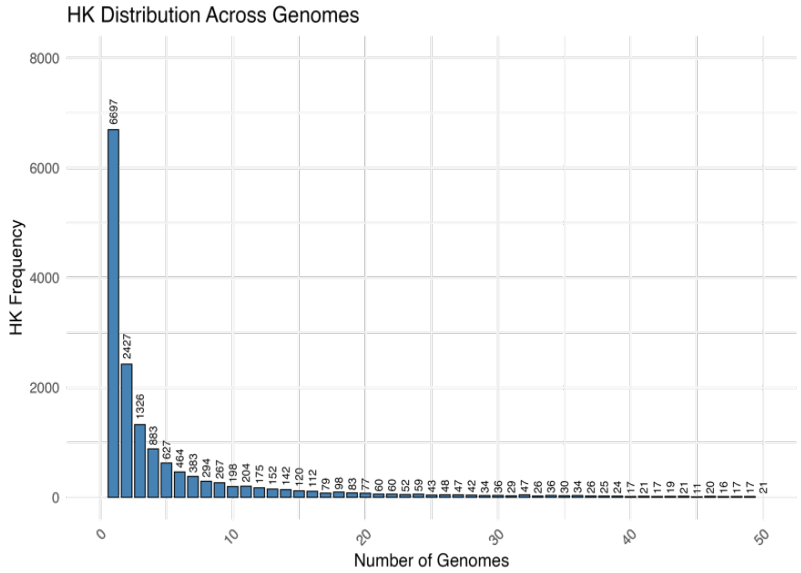


Figure 6 · HK Distribution across genomes, x-axis shows the number of genomes where a given number of HKs appeared, and y-axis shows numbers of unique HKs. One can see that the large majority of HKs were detected in fewer than 5 genomes.

For a better view, we examined Figure 7 where each genome had the total number of HKs annotated (dark blue) and, on top, with a light blue showing the HKs unique to that genome. In this graph, we reassured the findings on the other graph about the higher quantity of unique HKs and how most of the genomes contained more unique than common ones.

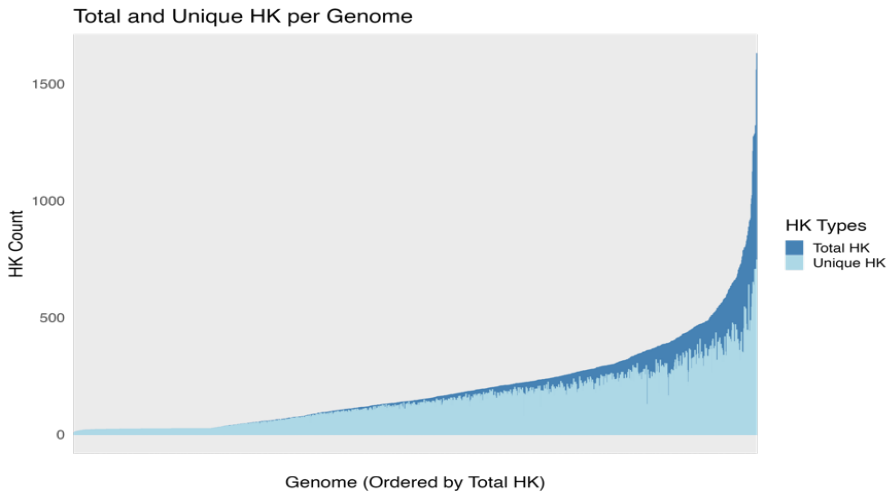


Figure 7· Distribution of total and unique HKs across genomes

Following the examination of the HKs on the different genomes, we saw a peak of 94 genomes with 25 HKs, and some previous bins above the others showed a great variability in the number of HKs for genomes in a rank from 5 to 300 HKs for genome. We also observed some other peaks slightly higher than the average with 12 and 13 genomes; however, we noted that many HKs appeared in only 3-7 genomes (Figure 8). The mean number of HKs per genome was 198 represented by a red line in the figure. We viewed the full distribution in Supplementary Figure 2, which showed that most genomes had a low number of HKs. There was a long flat tail extending to the right, where a small number of genomes contained a high HKs count.

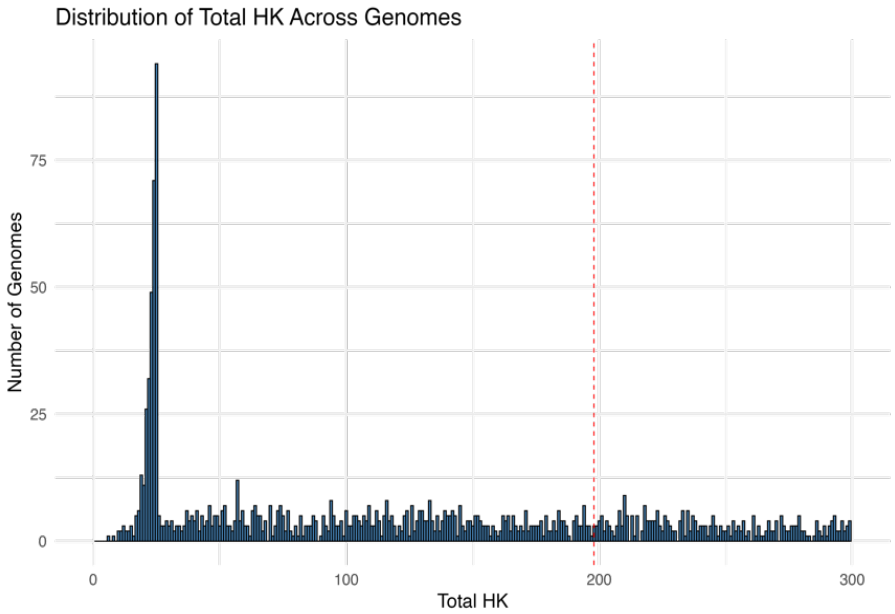


Figure 8· Distribution of total HKs across genomes

4.2. Correlation Analysis

The correlation between HKs length and the number of genomes where they appeared was examined (Figure 9). Initially, we expected that longer sequences would be more prevalent across multiple genomes because these would have been more likely to be annotated. However, the results indicated that long HKs were found in a limited number of genomes, and the huge majority were sequences of fewer than 200 nucleotides. A few HKs appeared in numerous genomes but had a sequence below 200 nucleotides, with only two exceptions that exceeded this threshold.

Figure 9 · Correlation between HK length and number of genomes

Another correlation analysis was conducted on the total number of HKs annotated and genome size, expecting that larger genomes would contain a larger number of HKs (Figure 10). As expected, genome size had a positive correlation with the number of HKs, we have a p-value < 0.0001. This trend aligned with the general principle that organisms with larger genomes often had more regulatory and functional elements, including HKs. We observed a clear overall increase in the total number of HKs as genome size expanded, however, the correlation was not particularly strong. Many large genomes also contained few HKs, indicating variability in HKs distribution.

This suggested that genome size alone did not dictate HKs abundance, implying that evolutionary pressure influencing genome expansion and HKs presence may have been independent or influenced by additional factors such as taxonomic affiliation beyond genome size alone. Some studies provided evidence of detailed statistics on the distribution of TCS proteins across genomes of varying sizes, and concluded that while larger genomes tended to have more HKs, the relationship was not linear, and there is substantial variability due to ecological and taxonomic factors. (Ulrich & Zhulin, 2010)

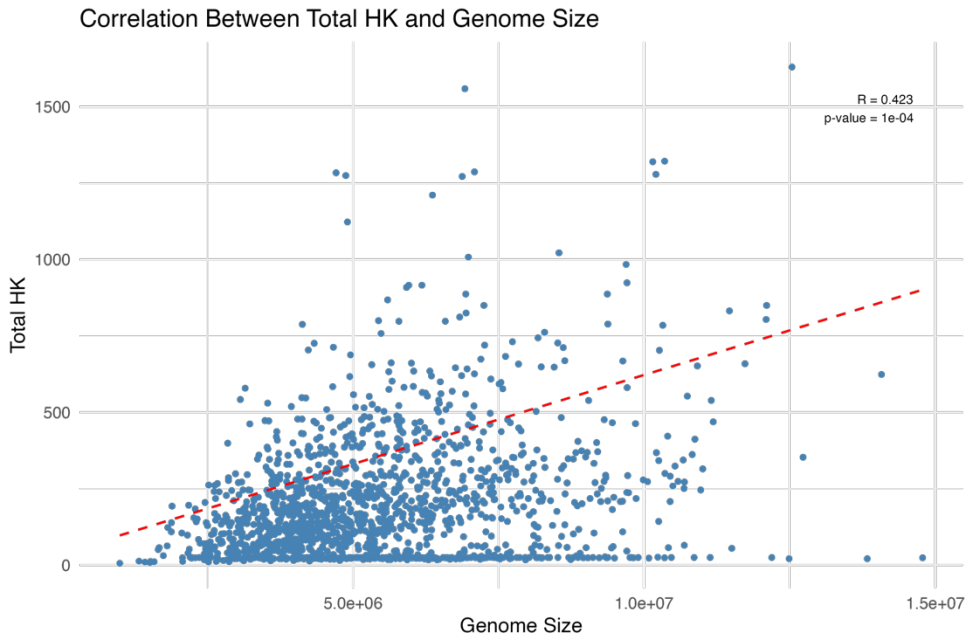


Figure 10 · Correlation between total HKs and genome size

4.3. HK by Response Group

To analyze the association between the composition of HKs and bacterial drought responsiveness, we examined their distribution using a Venn Diagram (Figure 11), which showed how they were distributed. We saw that the majority were shared between at least two genomes from the two groups (7,478) and a similar number were specific for each group: 4,974 were responsive and 4,149 were sensitive. We then repeated the comparison removing the HKs that were found in less than five genomes (Figure 12). This confirmed that 31% of the HKs were in five or more genomes, having a low minority being responsive or sensitive. These results showed that 7,478 HKs were shared between the responsive and sensitive groups, whereas only 4,974 were exclusive to responsive genomes and 4,149 to sensitive ones. However, this initial analysis included HKs that occurred in only a single genome, so we refined it by excluding those found in fewer than five genomes, which were considered rare (Figure 12). After this filtering, we observed that 260 HKs were specific to responsive genomes and 49 to sensitive ones, confirming that the proportion of distinguishing HKs was relatively low.

Interestingly, the number of HKs in responsive genomes was more than five times higher than in sensitive genomes. This could have been because drought resistance requireed specific adaptive mechanisms, whereas sensitivity to drought did not necessarily demand specialized responses. Another possible explanation was that the dataset contained more responsive genomes than sensitive ones, increasing the probability of identifying HKs

exclusively in the responsive group. Some researchers found that bacteria from variable stressful environments tended to have more TCS gens (Alm et al., 2006).

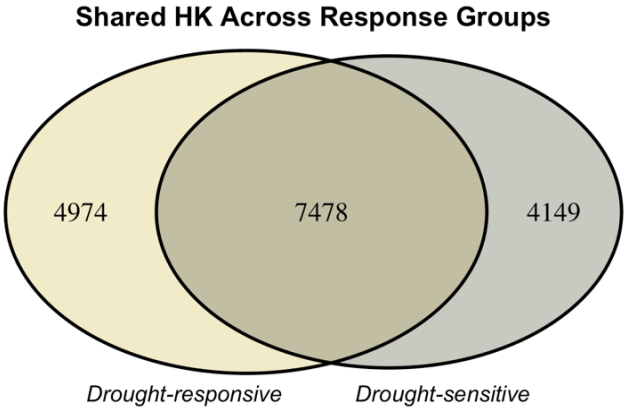


Figure 11 · Shared HKs across response groups

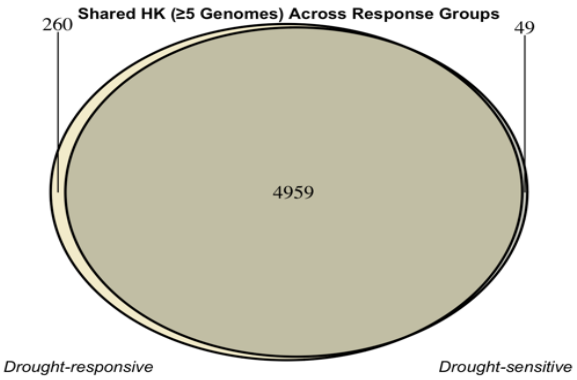


Figure 12 · Shared HKs in more than five genomes across response groups

A glance at the HKs data across phyla (Figure 13) revealed that 23.3% of the analyzed HKs were found in Proteobacteria, 12.4% in Actinobacteriota, and 11.2% in Acidobacteriota. These were followed by Myxococcota, Verrucomicrobiota, Bacteroidota, Firmicutes, Chloroflexi, and Planctomycetota, all of which decreased progressively but remained above 3%. Compared to the genome graph (Figure 3), the same top nine phyla appeared but in a different order, while Proteobacteria and Actinobacteriota remained the most represented.

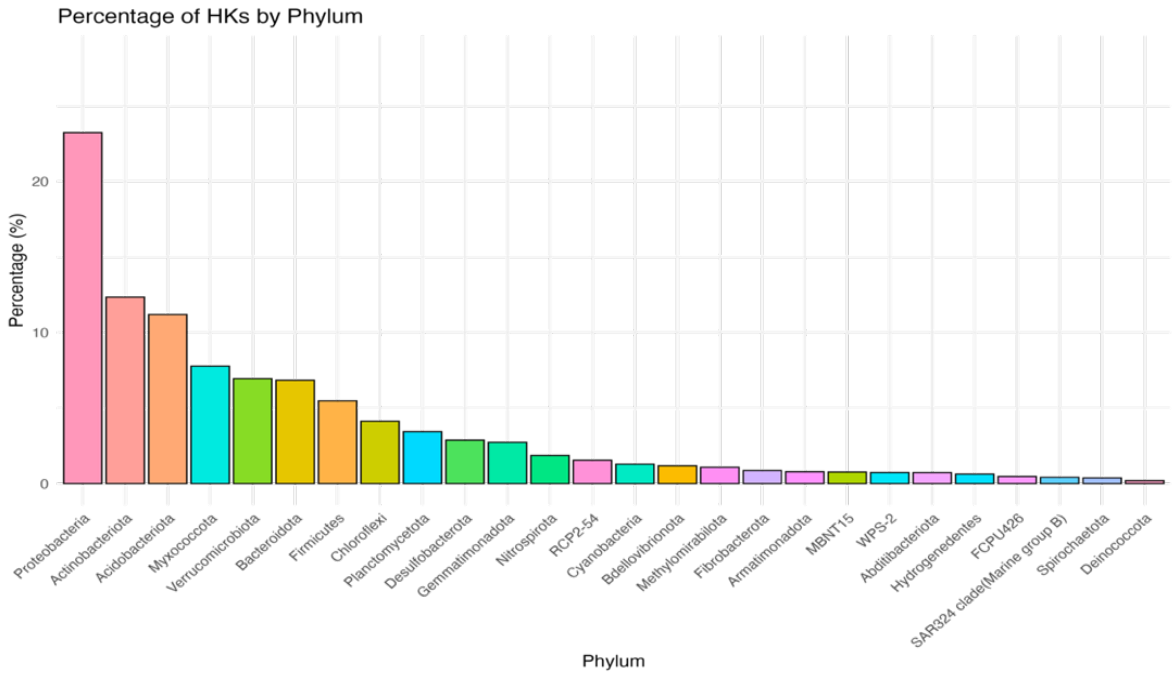


Figure 13 · Percentage of the total number of HKs annotated in the study found in each phylum.

On a further approach, we used the same distribution of HKs by phyla but separated it by response group (Figure 14). This data was proportional to the number of genomes in each phyla. Most of the phyla had a close equality share of both response groups with less of a 1% difference. In contrast, drought-responsive Actinobacteria had 8.5% of the total HKs and drought-sensitive a lower 3.6%, drought-responsive Firmicutes had 3.6 % and drought-sensitive a lower 2.5%.

Some investigations found how drought and other environmental changes affected the microbial community composition. They highlighted that Actinobacteria and Firmicutes often increased in abundance and functional gene content under drought, while other groups were less responsive or even declined (Martiny et al., 2017)(Evans & Wallenstein, 2012).We also identified some phyla with a low representation of less than 1% from only one response group, for drought-responsive Fibrobacteriota, WPs-2, and Hydrogenedentes. As for the drought-sensitive we found Methylomirabilota, Armatimonadota, MBNT15, FCPU426, SAR324 clade, Spirochaetota, and Deinococcota; it also has the phyla RCP2-54 had slightly more representation, 1.2%.

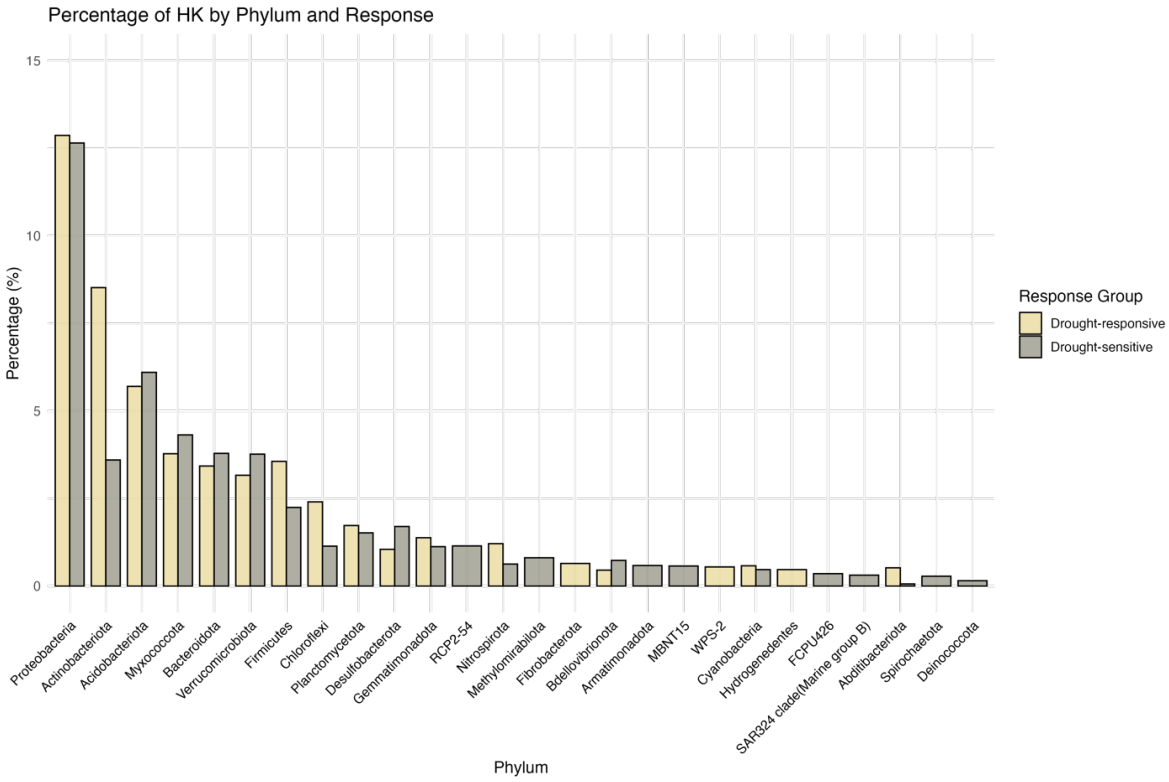


Figure 14 · Percentage of HKs by phylum and response

On a further visualization, we added the HKs counts per genome, removing in this way the influence of having more genomes of some phyla than others (Figure 15). We saww that there was a broad range and higher numbers of HKs per genome on some phyla like Proteobacteria, while others had lower counts like Bacteroidota. We observed a slightly higher diversity in HKs content among the genomes in the drought-responsive groups.

Further insight showed some phyla that could not be settled on a response group as there was no clear difference between the groups. There were some outliers on several phyla that had an exceptionally high HKs count, indicating that some individual genomes encoded for more HKs than the typical member of their phylum. This indicates that there is no direct relation between the HKs count and the drought response, making this more significant or relevant only in some bacterial groups.

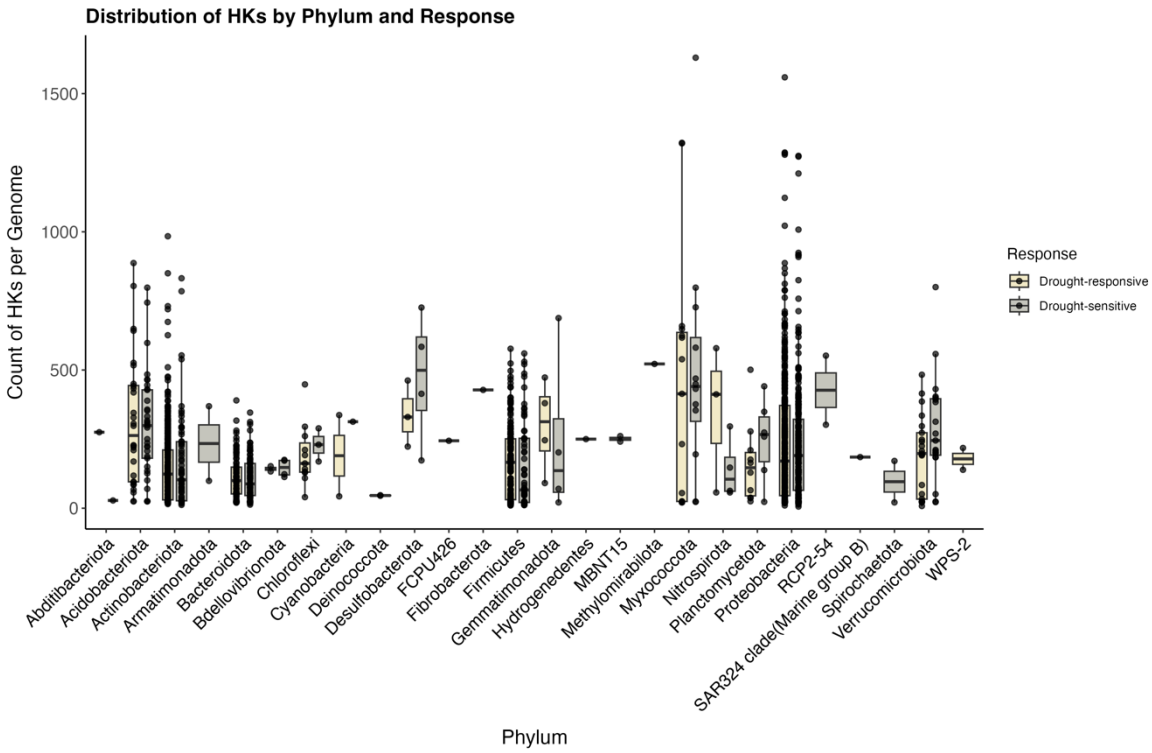


Figure 15 · Distribution of HKs by phylum and drought response

The correlation between the percentage of genomes and the percentage of HKs across the different phyla was represented by colored dots (Figure 16). This analysis had a strong correlation coefficient R = 0.8779 and a high significant p-value of 3.827e-09, showing a strong relation on the percentages of HKs and percentages of genomes from these phyla in the dataset. The x-axis position indicated its representation in the genome dataset, and the y-axis position reflected its contribution to the total HKs pool. We saw groups on the low part of the graph that had low percentages on both axes, these phyla have low contributions to the HKs pool; while we have one phyla Proteobacteria stood in the upper right, indicating substantial contributions to both categories.

This correlation confirmed the relation on the phyla representation increasing with the HKs counts proportionally, as demonstrated by the dashed trend line on the graph. Some phyla appeared above the trend line like Acidobacteriota, meaning that they possessed a higher percentage of HKs than expected for the number of genomes these contributed to the dataset; on the other hand, some phyla that below the trend line had proportionally fewer HKs. All these suggested that HKs generally had a distribution according to the genome representation across the phyla; also, some of the derivations above the trend line may have had evolutionary adaptations or ecological niches requiring a more extensive signaling network. The different clustering of the phyla suggested that HKs content had taxonomic patterns related to this differential investment in signal components. Some

reviews discussed the evolutionary and ecological factors influencing the abundance of TCS proteins, it highlighted that there was a general scaling relationship, but deviations occured due to lineage-specific adaptations and ecological pressures, resulting in some phyla investing more or less in signaling networks (Wuichet et al., 2010).

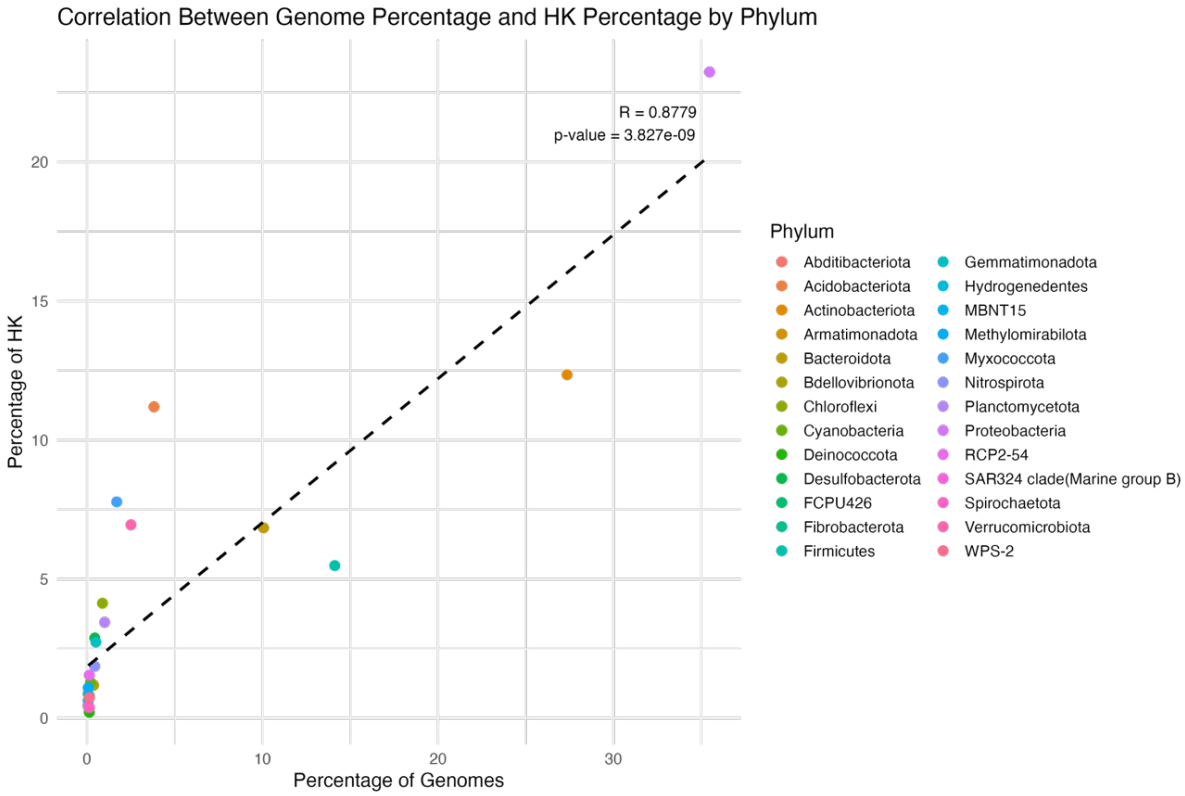


Figure 16 · Correlation between genome percentatge (% of genomes belonging to a given phylum) and HKs percentatge (% of HKs found in genomes of these phyla).

This last boxplot (Figure 17) revealed the distribution of HKs for megabase across bacterial phyla. The x-axis listed different bacterial phyla, while the y-axis showed the number of histidine kinases (HKs) per megabase pair (Mbp) of the genome. A comparison of HKs density and the variability within and between bacterial phyla was made. This revealed substantial differences in HKs gene density among bacterial phyla. Some phyla, such as Fibrobacterota and Chlorobiota, had notably high median HK densities (around 80–90 HKs/Mbp), suggesting a strong investment in environmental signal transduction. Proteobacteria stood out for both their high median density and exceptional variability, with some genomes reaching nearly 300 HKs/Mbp, indicating diverse ecological strategies even within a single phylum. This matched with some articles that investigated the link of Proteobacteria to diverse ecological strategies and metabolic capacities,

particularly in hydrothermal environments, underscoring their adaptability and niche specificity (Zhou et al., 2020).

In contrast, phyla like Bacteroidota, Acidobacteriota, and Actinobacteriota showed consistently lower HK densities (20–40 HKs/Mbp), implying more streamlined signaling systems. Bacteoidetes may have utilized signaling systems similar to the two components, enhancing their communication and interaction within microbial communities (Blackwell & Fuqua, 2011) typically associated with simpler one-component systems (Seshasayee & Luscombe, 2011). Some studies indicated that acidophiles exhibited a lower representation of genes encoding signal transduction mechanisms. This was caused caused by the energy optimization to challenge low-pH environments that they inhabited (and slower growth rates, emphasizing their smaller genomes (Cortez et al., 2022). Actinobacteria typically hd fewer TCS components, reflecting more streamlined signaling networks (M. Y. Galperin, 2018).

The SAR324 clade displayed a high but homogenous HK density, suggesting a specialized but uniform adaptation. One study analyzed all publicly available SAR324 genomes and reconstructed additional genomes from diverse marine environments. This paper found that SAR324 displays a broad metabolic potential but also noted that, within specific clades, there is a relatively uniform set of genes involved in environmental sensing and adaptation (Malfertheiner et al., 2022). The presence of outliers and wide variation within some phyla, especially Proteobacteria, highlighted the adaptive flexibility of bacterial signaling networks in response to specific environmental pressures. Overall, the data supported the idea that HK density was shaped by both phylogenetic lineage and ecological context, with some groups evolving extensive signaling networks in response to environmental complexity, while others maintained minimal systems suited to stable or specialized habitats. This aligned with recent comparative genomic studies (Capra & Laub, 2012), which highlighted the evolutionary flexibility and ecological significance of bacterial signaling architectures.

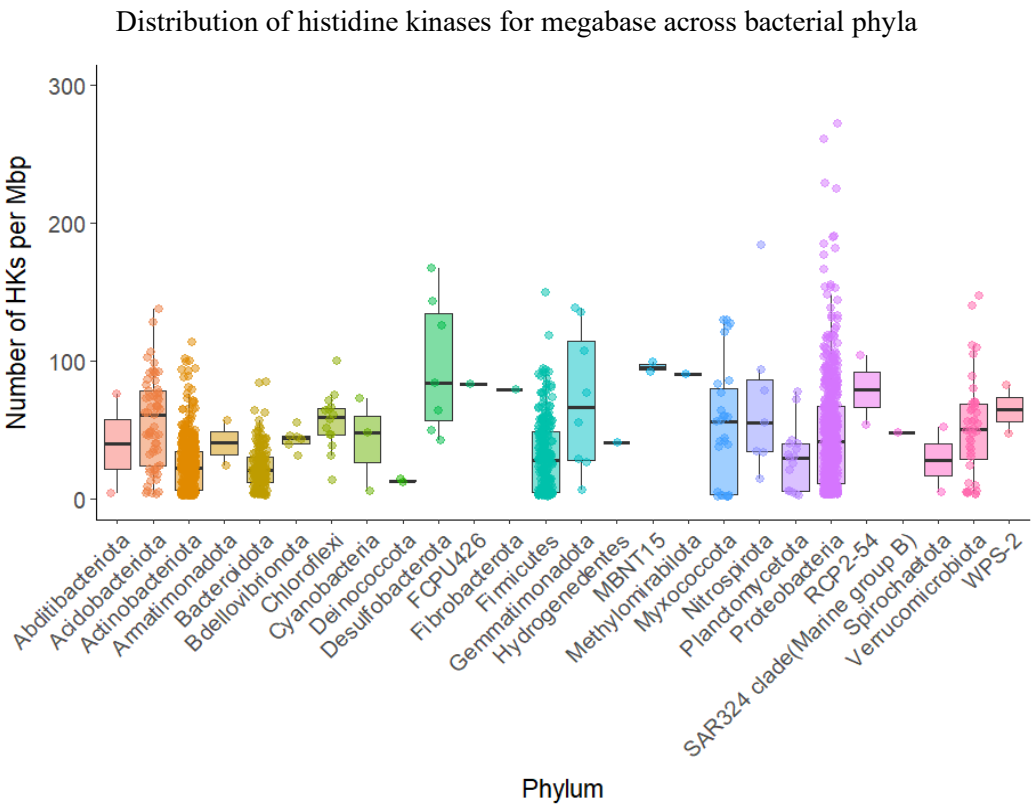


Figure 17 · Distribution of histidine kinases for megabase across bacterial phyla

5. Conclusions

- The majority of the HK are rare or functionally specialized, they appear only in a few genomes, while a small number is widely conserved, likely reflecting essential biological roles.

- There is significant variability in the number of HKs per genome, ranging from 5 to 300, with most genomes containing more unique HKs than shared ones, highlighting the diversity of signaling strategies among bacteria.

- No significant correlation was found between HK sequence length and its prevalence across genomes. While larger genomes tend to have more HKs, the correlation is weak, indicating that factors beyond genome size influence HK content.

- Most HKs are shared between drought-responsive and drought-sensitive groups. However, after filtering for more prevalent HKs, only a small fraction is unique to either group, suggesting that most HKs are not exclusively associated with drought response. Notably, drought-responsive unique HKs are 5 times higher than drought-sensitive. .

- The distribution of HKs across phyla is uneven, with Proteobacteria, Actinobacteriota, and Acidobacteriota being the most represented. Some phyla show a higher proportion of HKs than expected based on their genome representation, suggesting lineage-specific adaptations.

- HK density varies dramatically between phyla, with some groups (e.g., Proteobacteria, Fibrobacterota, Chlorobiota, Desulfobacterota) exhibiting high median HK densities and others (e.g., Bacteroidota, Acidobacteriota, Actinobacteriota) showing lower values. This reflects different evolutionary pressures and ecological strategies, where higher HK densities might be linked to enhanced environmental sensing in complex or variable habitats.

- The observed patterns suggest that HK content is shaped by both phylogenetic heritage and ecological requirements, with some bacteria that might be investing more energy in signaling networks to adapt to diverse or challenging environments.

5.1. Limitations and Improvements

Some limitations of this work have been the limited number of HKs for drought-response. The identification and annotations of HK rely on sequence homology, this percentage can be modified to have a higher percentage. A larger number of data can be incremented to have a higher percentage of homology between sequences. These response HKs would be interesting for further investigation regarding aridity gradients with their phylogenetic and ecological distribution. We also have the question of why some genomes have so few HKs. One possibility is that they belong to rare species, and our HK database did not contain HKs similar to theirs. Alternatively, they could be parasites or have adaptations that make responding to environmental changes less crucial. This has space for further investigation.

Potential improvements are the inclusion of more genomes from underrepresented phyla and rare HKs, metagenome-assembled genomes (MAGs), or environmental samples to capture broader HK diversity. Updating the dataset with more genomes to increase the information available. Incorporate environmental metadata (e.g., habitat type, climate data) to better correlate HK content with the ecological context of the bacterial taxa. Use more precise or experimentally validated criteria for grouping genomes (e.g., confirmed drought tolerance assays). Complement bioinformatic predictions with laboratory experiments to test the functional roles of specific HKs in environmental sensing and adaptation. Analyze the function and ecological significance of rare HKs or those found only in specific lineages, as these may represent novel adaptations.

6. Bibliografia

Alm, E., Huang, K., & Arkin, A. (2006). The Evolution of Two-Component Systems in Bacteria Reveals Different Strategies for Niche Adaptation. *PLoS Computational Biology*, *2*(11), e143. https://doi.org/10.1371/journal.pcbi.0020143

Bhate, M. P., Molnar, K. S., Goulian, M., & DeGrado, W. F. (2015). Signal Transduction in Histidine Kinases: Insights from New Structures. *Structure*, *23*(6), 981–994. https://doi.org/10.1016/j.str.2015.04.002

Blackwell, H. E., & Fuqua, C. (2011). Introduction to Bacterial Signals and Chemical Communication. *Chemical Reviews*, *111*(1), 1–3. https://doi.org/10.1021/cr100407j

Capra, E. J., & Laub, M. T. (2012). Evolution of two-component signal transduction systems. *Annual Review of Microbiology*, *66*, 325–347. https://doi.org/10.1146/annurev-micro-092611-150039

Cortez, D., Neira, G., González, C., Vergara, E., & Holmes, D. S. (2022). A Large-Scale Genome-Based Survey of Acidophilic Bacteria Suggests That Genome Streamlining Is an Adaption for Life at Low pH. *Frontiers in Microbiology*, *13*, 803241. https://doi.org/10.3389/fmicb.2022.803241

Evans, S. E., & Wallenstein, M. D. (2012). Soil microbial community response to drying and rewetting stress: Does historical precipitation regime matter? *Biogeochemistry*, *109*(1/3), 101–116.

Feng, S., & Fu, Q. (2013). Expansion of global drylands under a warming climate. *Atmospheric Chemistry and Physics*, *13*(19), 10081–10094. https://doi.org/10.5194/acp-13-10081-2013

Foster, J., Sheng, Q., McClain, J., Bures, M., Nicas, T., Henry, K., Winkler, M., & Gilmour, R. (2004). Kinetic and mechanistic analyses of new classes of inhibitors of two-component signal transduction systems using a coupled assay containing HpkA-DrrA from Thermotoga maritima. *Microbiology (Reading, England)*, *150*, 885–896.

Galperin, M. (2005). A census of membrane-bound and intracellular signal transduction proteins in bacteria: Bacterial IQ, extroverts and introverts. *BMC Microbiology*, *5*, 35. https://doi.org/10.1186/1471-2180-5-35

Galperin, M. Y. (2018). What bacteria want. *Environmental Microbiology*, *20*(12), 4221–4229. https://doi.org/10.1111/1462-2920.14398

Louca, S., Parfrey, L. W., & Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. *Science*, *353*(6305), 1272–1277. https://doi.org/10.1126/science.aaf4507

Malfertheiner, L., Martínez-Pérez, C., Zhao, Z., Herndl, G. J., & Baltar, F. (2022). Phylogeny and Metabolic Potential of the Candidate Phylum SAR324. *Biology*, *11*(4), 599. https://doi.org/10.3390/biology11040599

Malik, A. A., & Bouskill, N. J. (2022). Drought impacts on microbial trait distribution and feedback to soil carbon cycling. *Functional Ecology*, *36*(6), 1442–1456. https://doi.org/10.1111/1365-2435.14010

Martiny, J. B., Martiny, A. C., Weihe, C., Lu, Y., Berlemont, R., Brodie, E. L., Goulden, M. L., Treseder, K. K., & Allison, S. D. (2017). Microbial legacies alter decomposition in response to simulated global change. *The ISME Journal*, *11*(2), 490–499. https://doi.org/10.1038/ismej.2016.122

Philippot, L., Griffiths, B. S., & Langenheder, S. (2021). Microbial Community Resilience across Ecosystems and Multiple Disturbances. *Microbiology and Molecular Biology Reviews : MMBR*, *85*(2), e00026-20. https://doi.org/10.1128/MMBR.00026-20

Schimel, J. P. (2018). Life in Dry Soils: Effects of Drought on Soil Microbial Communities and Processes. *Annual Review of Ecology, Evolution, and Systematics*, *49*(1), 409–432. https://doi.org/10.1146/annurev-ecolsys-110617-062614

Schulz, S., Brankatschk, R., Dümig, A., Kögel-Knabner, I., Schloter, M., & Zeyer, J. (2013). The role of microorganisms at different stages of ecosystem development for soil formation. *Biogeosciences*, *10*(6), 3983–3996. https://doi.org/10.5194/bg-10-3983-2013

Seshasayee, A. S. N., & Luscombe, N. M. (2011). Comparative genomics suggests differential deployment of linear and branched signaling across bacteria. *Molecular BioSystems*, *7*(11), 3042. https://doi.org/10.1039/c1mb05260h

Tariq, A., Sardans, J., Zeng, F., Graciano, C., Hughes, A. C., Farré-Armengol, G., & Peñuelas, J. (2024). Impact of aridity rise and arid lands expansion on carbon-storing capacity, biodiversity loss, and ecosystem services. *Global Change Biology*, *30*(4), e17292. https://doi.org/10.1111/gcb.17292

Tran, L.-S. P., Urao, T., Qin, F., Maruyama, K., Kakimoto, T., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2007). Functional analysis of AHK1/ATHK1 and cytokinin receptor histidine kinases in response to abscisic acid, drought, and salt stress in Arabidopsis. *Proceedings of the National Academy of Sciences*, *104*(51), 20623–20628. https://doi.org/10.1073/pnas.0706547105

Ulrich, L. E., & Zhulin, I. B. (2010). The MiST2 database: A comprehensive genomics resource on microbial signal transduction. *Nucleic Acids Research*, *38*(Database issue), D401-407. https://doi.org/10.1093/nar/gkp940

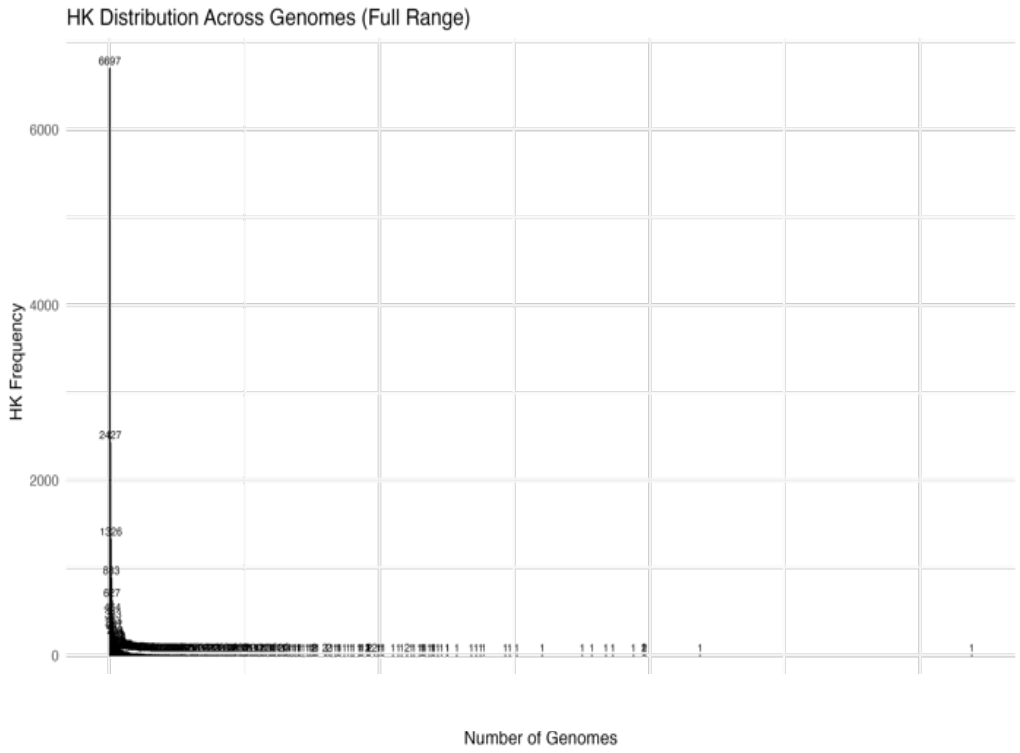
Wolanin, P. M., Thomason, P. A., & Stock, J. B. (2002). Histidine protein kinases: Key signal transducers outside the animal kingdom. *Genome Biology*, *3*(10), reviews3013.1. https://doi.org/10.1186/gb-2002-3-10-reviews3013

Wuichet, K., Cantwell, B. J., & Zhulin, I. B. (2010). Evolution and phyletic distribution of two-component signal transduction systems. *Current Opinion in Microbiology*, *13*(2), 219–225. https://doi.org/10.1016/j.mib.2009.12.011

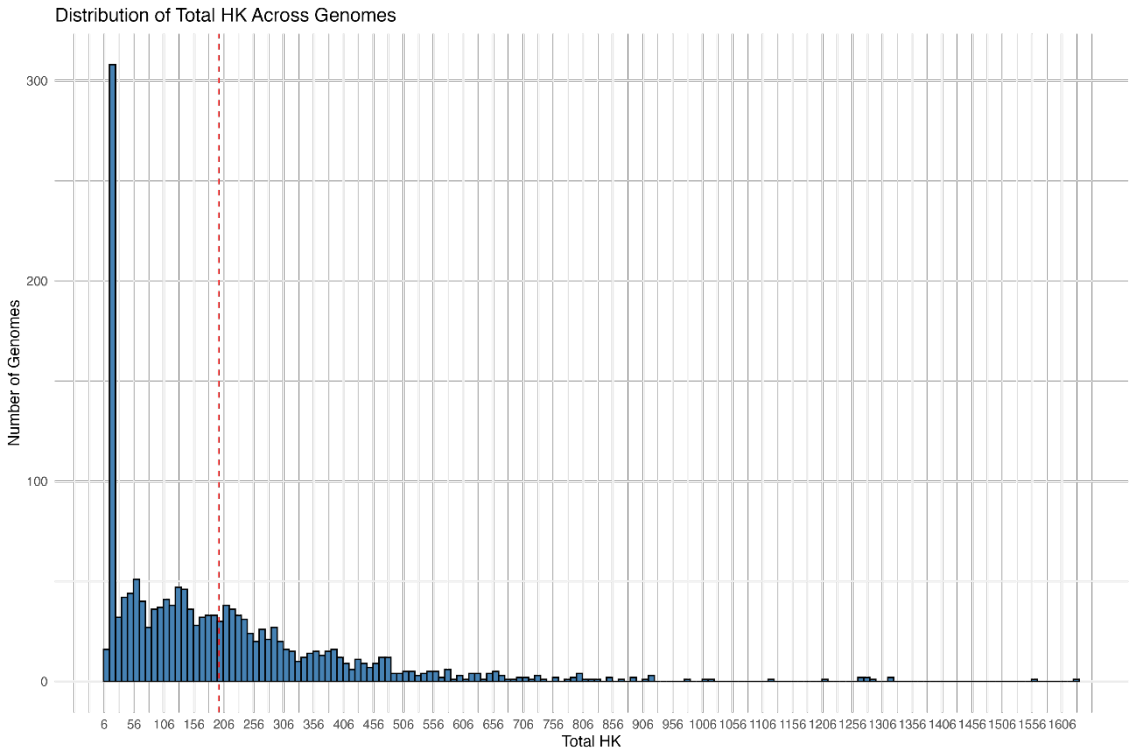
Zhou, Z., Tran, P. Q., Kieft, K., & Anantharaman, K. (2020). Genome diversification in globally distributed novel marine Proteobacteria is linked to environmental adaptation.

The ISME Journal, *14*(8), 2060–2077. https://doi.org/10.1038/s41396-020-0669-4

Annex A



Supplementary Figure 1 HK full Distribution across genomes



Supplementary Figure 2 Distribution of Total HK Across Genomes full length