CODON AND AMINO ACID USAGE ANALYSIS IN THE GENUS STAPHYLOCOCCUS

Thesis Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

By

Pinky

Registration Number: 42100102

Supervised By

Dr. Sandeep Kaur (28639)

Department of Medical Laboratory Sciences

(Assistant Professor)

School of Allied Medical Sciences

LOVELY PROFESSIONAL UNIVERSITY, PUNJAB

2025

DECLARATION

I, hereby declared that the presented work in the thesis entitled "**Codon And Amino Acid Usage Analysis In The Genus Staphylococcus**" in fulfillment of degree of **Doctor of Philosophy (Ph. D.)** is outcome of research work carried out by me under the supervision of Dr. Sandeep Kaur working as Assistant Professor, in the School of Allied Medical Sciences of Lovely Professional University, Punjab, India. In keeping with general practice of reporting scientific observations, due acknowledgments have been made whenever work described here has been based on findings of other investigator. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

(Signature of Scholar)

Name of the scholar: Pinky

Registration No.: 42100102

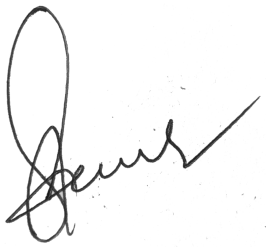
Department/school: School of Bioengineering and Biosciences

Lovely Professional University,

Punjab, India

CERTIFICATE

This is to certify that the work reported in the Ph. D. thesis entitled "**Codon And Amino Acid Usage Analysis In The Genus Staphylococcus**" submitted in fulfillment of the requirement for the award of the degree of **Doctor of Philosophy (Ph.D.)** in the School of Pharmaceutical Sciences, is a research work carried out by Pinky, 42100102, is bonafide record of his/her original work carried out under my supervision and that no part of thesis has been submitted for any other degree, diploma or equivalent course.



(Signature of Supervisor)

Name of supervisor: Dr. Sandeep Kaur

Designation: Assistant Professor

Department/school: Department of Medical Laboratory Sciences, School of Allied Medical Sciences

University: Lovely Professional University

ABSTRACT

The *Staphylococcus* genus, comprising Gram-positive bacteria, is of paramount importance in microbiology and medicine due to its role in human infections and the significant challenge it presents through antibiotic resistance. Prominent species such as *Staphylococcus aureus* and *Staphylococcus epidermidis* are implicated in a variety of diseases, ranging from superficial skin infections to severe systemic conditions. A deeper understanding of the genetic structure and codon usage patterns of *Staphylococcus* species can provide critical insights into their evolutionary adaptations, pathogenic traits, and mechanisms for survival within the host. In this study, we employed a comprehensive suite of analytical approaches to investigate codon usage across 48 species of the *Staphylococcus* genus. Key methodologies included the analysis of the effective number of codons (ENc), relative synonymous codon use (RSCU), and codon adaptation index (CAI), alongside neutrality plots, correspondence analysis (COA), and parity plot analysis to discern codon usage biases and their evolutionary implications. Our initial investigation into the genomic G-C content across these 48 *Staphylococcus* species revealed a notable variation, with values ranging from 34.27% to 36%. The coagulase-positive species (CoPS) were observed to have a higher G-C content, averaging 36%, compared to coagulase-negative species (CoNS) and co-variables, which exhibited lower G-C contents of 34.47% and 34.29%, respectively. These variations in G-C content significantly influenced codon usage patterns within the genus, contributing to adaptive traits that optimize gene expression. We observed a marked preference for A/T nucleotides at the third codon position, with CoPS showing a 72% preference, CoNS at 73.07%, and co-variables at 73.91%. This bias towards AT-rich codons is linked to the genomic structure and has profound implications for the adaptability of these species. A genome enriched in A/T content supports growth and sustenance post-infection by facilitating metabolic processes through increased availability of A/T-based metabolites. Additionally, AT-rich regions contribute to the efficiency of DNA unwinding and replication initiation, essential for rapid proliferation. The analysis of GC3s versus ENc revealed significant codon usage bias among CoPS, CoNS, and co-variables. A discernible trend in bias was identified, decreasing in the order of co-variables > CoPS > CoNS. Elevated ENc values across

these groups indicated a preference for synonymous codons with reduced bias, suggesting that translational selection plays a critical role in shaping codon usage. This preference aligns with the hypothesis that pathogens evolve to optimize their codon usage to minimize tRNA competition, thereby streamlining gene expression. Notably, CoPS showed a reduced gene count compared to CoNS and co-variables, hinting at an adaptation that balances translational selection and mutational pressure. These findings underscore the nuanced interplay between evolutionary pressures and genome optimization within the *Staphylococcus* genus. The neutrality plot between GC3s and GC12 provided further insights into the extent of mutational pressure versus natural selection. Negative regression coefficients close to zero for both CoPS (-0.015) and CoNS (-0.026) underscored the predominance of natural or translational selection over mutational constraints. A positive coefficient for coagulase-variables (0.05) similarly pointed to the influence of selection in shaping codon usage. The distinct clustering of CoPS, CoNS, and co-variables on genomic data plots reinforced the hypothesis that evolutionary processes drive codon usage, with pathogenic strains showing unique adaptations in their genetic makeup. These clustering patterns are consistent with shifts towards pathogenicity, supporting the idea that natural selection significantly impacts the genomic evolution of these organisms. Parity plot analysis (PR2) provided additional depth, shedding light on nucleotide composition dynamics. By examining the relationship between G3s/G3s + C3s and A3s/A3s + T3s, we identified a preference for purines over pyrimidines at the third codon position, with mean values across the *Staphylococcus* genus above 0.5. This trend highlighted a bias towards A and G nucleotides, which may confer adaptive benefits by enhancing the efficiency of protein synthesis and reducing the metabolic burden. These findings align with the broader genomic architecture that favors AT-rich codons, influencing replication and transcription processes. Dinucleotide pair analysis further revealed the expression patterns critical to understanding pathogenic potential. CoPS exhibited a higher frequency of GpC, ApC, GpA, and TpG dinucleotide pairs, whereas GpG, CpC, and CpT were least expressed. In CoNS, GpC, ApC, and GpA were similarly prevalent, with CpC, GpG, and TpC being the least expressed. The low abundance of CpG, a dinucleotide known to trigger immune responses, is particularly significant as it may contribute to immune evasion strategies. This underrepresentation is consistent with the

adaptive evolution of pathogens to reduce immune detection and enhance survival within the host. Codon usage bias was also evident, with TTA (Leu) being the most favored codon across the genus. Specific biases were observed in CoPS, where CGT (Arg), GGT (Gly), and AGA (Arg) were predominant, while CCC (Pro), TCC (Ser), and CTG (Leu) were less favored. This preference for AT-rich codons, particularly TTA, reflects evolutionary pressures that favor codons associated with energy-efficient amino acid usage, enhancing survivability in host environments. The exclusive use of TTA for leucine throughout the genus emphasizes its critical role in adaptive evolution. The analysis extended to codon pairs and dinucleotide expressions at junctions, where NNG-CNN and NNA-CNN codon pairs were overrepresented, aligning with the prominence of GpC and ApC dinucleotide pairs. In contrast, NNT-CNN and NNG-GNN were underrepresented, indicating reduced TpC and GpG presence. These patterns are congruent with dinucleotide trends observed, highlighting the interplay between codon and dinucleotide biases in shaping the genetic architecture. Our study found a notable correlation between CAI and G-C content, with higher CAI observed in CoNS, indicating significant bias in synonymous codon usage. This correlation, alongside relationships with RSCU Axis1 and Axis2, underscores the complex relationship between genomic composition and gene expression. Amino acid usage analysis across the genus revealed a preference for leucine, isoleucine, and lysine, whereas cysteine and methionine were underrepresented. This trend suggests a selection for amino acids that promote energy efficiency. GRAVY and AROMO analyses highlighted correlations between amino acid hydropathicity and relative usage, influenced by genomic GC content.

These insights underscore the evolutionary strategies of *Staphylococcus* species, revealing their capacity to adapt genomic codon usage for optimized gene expression and survival within diverse environments. Such adaptations have significant implications for understanding the pathogenicity, resistance mechanisms, and evolutionary biology of these clinically significant bacteria.

Keywords: *Staphylococcus*, codon usage bias, evolutionary adaptation, genomic G-C content, translational selection, dinucleotide analysis.

ACKNOWLEDGEMENT

It is a moment of gratification and pride to look back with a sense of contentment at the long-travelled path, to be able to recapture some of the fine moments, to be able to think of the infinite number of people, some who were with me from the beginning, some who joined me at some stage during the journey, whose kindness, love and blessings has brought me to this day. I wish to thank each one of them from the bottom of my heart First and foremost, I would like to thanks the "**almighty**" for showering blessings of life and wisdom over me and for never forsaking me, even when I have forgotten, at too many occasions, to pray and thank him for all the blessings that he bestows on me. I express my thanks to the Chancellor sir, **Dr. AshokMittal** and Pro-Chancellor Mam **Dr. Rashmi Mittal** for providing us supportive and friendly atmosphere, excellent research facilities in and around the region, exposure to the scientific word and platform to rise. I would like to acknowledge my indebtedness and render my warmest thanks to my supervisor, **Dr. Sandeep Kaur** who made this work possible. Their guidance and expert advice have been invaluable throughout all stages of the work. I appreciate all his contributions of time, ideas, and efforts to make my Ph.D experience productive and stimulating and imbibed the strength in me to work hard through this endeavour. Indeed, the words at my command are inadequate to express my gratifications to my esteemed teacher and guide who helped me in struggling this path of research. A journey is easier when youtravel together. I would like to thank **Dr. Shelly Gupta and Dr. Imran** for thier guidance in the initial tenure of my Ph.D. I would like to show my greatest appreciation to for their valuable suggestions and approval of my research work andexemplary recognition. I would also like to thanks Ph.D. scholars and my friends **Dr. Anuj Sharma** and **Mr. Ujwal Dahal** for their time to time help,support, guidance. I feel highly obliged and have immense pleasure in expressing my heartfelt thanks to Dean **Prof. Neeta Raj Sharma** as well as HOD **Dr. Ashish Vyas**, **Dr. Arvind Kumar**, **Dr. Minhaj** and **Dr Jeena Gupta.** I extend my sincere greetings and gratitude to **Dr. C. S. Mukhopadhyay, Senior Scientist**, **Department of Bioinformatics, College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University**, whose invaluable guidance has been instrumental in bringing me to this point. Thank you very much, sir, for

nurturing my future. I expressmy sincere thanks to my lovable siblings and cousins **Ms. Dinky Munjal**, **Mr. Kunal Arora, Ms. Reena Rani and Mr. Dheeraj Kumar.** This work would not have been possible without their, support and encouragement. I would also like to thank my colleagues and friends **Ms. Anshika Sharma, Dr. Anuradha, Ms. Simranjeet Kaur, and Ms. Navneet Kaur** for their continuous support. Last but not the least, this thesiswould not have been possible without the confidence, endurance and support of my family. My family has always been a source of inspiration and encouragement. I wish to thanks my father Sh. **Madan Lal**, mother **Smt. Neelam**. Last but the most important person who has always guided me during my Ph.D, supported me in my tough times, and always cared for me in every situation, **Dr. Shubham Kumar**, and this work would not be possible without your encouragement, love, kind support and believe. Thank you, my dear love, for always being on my side and helped me in every situation. Above all I thank **Lord Hanuman** and My parents for showering their infinite bounties, clemencies and graces upon me and for being my constants companions, the strongest source of motivation, inspiration and my ultimate Guardians; to them I owe a lifelong indebtedness.

Pinky Arora

Dedicated to

My Family

TABLE OF CONTENTS

1.	**Introduction**			**1-22**
	1.1.	Background		1-2
	1.2.	Isolation and History		2-3
	1.3.	Taxonomic History		3-4
	1.4.	Pathogenesis		4
	1.5.	Coagulase Positive Staphylococci		4-7
		1.5.1.	*Staphylococcus aureus*	5
		1.5.2.	*Staphylococcus intermedius*	5-6
		1.5.3.	*Staphylococcus pseudintermedius*	6
		1.5.4.	*Staphylococcus delphini*	6
		1.5.5.	*Staphylococcus argenteus*	6
		1.5.6.	*Staphylococcus schleiferi subsp. coagulans*	7
		1.5.7.	*Staphylococcus agnetis*	7
		1.5.8.	*Staphylococcus hyicus*	7
	1.6.	Coagulase Negative Staphylococci		8-17
		1.6.1.	*Staphylococcus arlettae*	9-10
		1.6.2.	*Staphylococcus aureus*	10
		1.6.3.	*Staphylococcus aureus subsp. anaerobius*	10
		1.6.4.	*Staphylococcus capitis*	11
		1.6.5.	*Staphylococcus caprae*	11
		1.6.6.	*Staphylococcus carnosus*	11-12
		1.6.7.	*Staphylococcus cohnii*	12
		1.6.8.	*Staphylococcus delphini*	12
		1.6.9.	*Staphylococcus epidermidis*	12-13
		1.6.10.	*Staphylococcus equorum*	13
		1.6.11.	*Staphylococcus gallinarum*	13
		1.6.12.	*Staphylococcus haemolyticus*	13-14
		1.6.13.	*Staphylococcus hominis*	14
		1.6.14.	*Staphylococcus intermedius*	14-15
		1.6.15.	*Staphylococcus kloosii*	15
		1.6.16.	*Staphylococcus lentus*	15
		1.6.17.	*Staphylococcus saprophyticus*	15
		1.6.18.	*Staphylococcus sciuri*	15-16

		1.6.19.	*Staphylococcus simulans*	16
		1.6.20.	*Staphylococcus vitulus*	16
		1.6.21.	*Staphylococcus warneri*	16-17
		1.6.22.	*Staphylococcus xylosus*	17
	1.7.	Codon Usage Bias		17
		1.7.1.	*Codon Usage in Staphylococcus: Overview*	17-18
		1.7.2.	*Influences on Codon Usage Bias*	18
	1.8.	Amino Acid Usage Analysis		18
	1.9.	Mechanisms Behind Codon Usage Bias		19
	1.10.	Importance of Codon and Amino Acid Analysis in *Staphylococcus*		19-20
	1.11.	Rationale for Studying Codon and Amino Acid Usage in *Staphylococcus*		20-22
2.	**Review of Literature**			**23-40**
	2.1.	Codon Usage Bias Across Domains of Life		23-27
	2.2.	Codon Usage Bias in Plants		27-31
	2.3.	Codon Usage Bias in Fungal Systems		31-33
	2.4.	Functional and Mechanistic Insights into Codon Usage Bias		33-37
	2.5.	Codon Usage Bias in Viruses		37-40
3.	**Research Gap, Aim and Objective**			**41-42**
	3.1.	Research Gap		41
	3.2.	Aim		42
	3.3.	Objectives		42
4.	**Materials and Methods**			**43-50**
	4.1	Data collection and genome annotation		43
	4.2.	Codon usage analysis tools and statistical measures		43-45
	4.3	Neutrality plot		46
	4.4	Assessment of translational selection		46-47
	4.5.	Codon adaptation index		47
	4.6.	Correspondence analysis		47
	4.7.	Analysis of relative abundance of dinucleotides		48
	4.8.	Estimation of relative synonymous codon pair usage and codon pair score		48
	4.9.	Statistical analysis		49
	4.10.	Experimental validation of codon usage bias in selected species by genome sequencing		49-50
		4.10.1.	*Preparation of Cultures*	49

		4.10.2.	*DNA Extraction*	50
		4.10.3.	*Genome Sequencing*	50
5.	**Results and Discussion**			**51-165**
	5.1.	AT-rich *Staphylococcal* genomes elegantly shape the selection of preferred codons		51-57
	5.2.	ENc influencing codon usage bias		57-64
	5.3.	Neutrality Plot: Deciphering Selection Forces in *Staphylococcus* Genome		65-72
	5.4.	Preferential Purine Utilization in *Staphylococcus* Genomes		73-80
	5.5.	Distinct Dinucleotide Patterns in *Staphylococcus* Variants		81-89
	5.6.	Codon Preferences in the Genomic Landscape		89-95
	5.7.	Exploring Codon Context Variations at Junctions		95-96
	5.8.	Exploring Synonymous Codon Similarity in *Staphylococcal* Genomes		96-145
	5.9.	Pathogens tends to favor amino acid with low biosynthetic costs		146-149
	5.10.	Hydrophobicity and Aromaticity shape amino acid usage		149-158
	5.11.	Corresponding Analysis differentiates between CoPS, CoNS and Co variables		158-160
	5.12.	Validation of codon usage bias in selected species by genome sequencing		160-161
	5.13.	Discussion		162-165
6.	**Conclusion**			**166-170**
7.	**References**			**171-185**

LIST OF FIGURES

Figure Number	Title	Page Number
Figure 1	Classification of Genus *Staphylococcus*	8
Figure 2A	GC3 *vs* Enc plot, with each data point representing an individual gene A). *Staphylococcus argenteus str. 58113* B). *Staphylococcus arlettae* C). *Staphylococcus auricularis* D). *Staphylococcus capitis* E). *Staphylococcus caprae M23864:W1* F). *Staphylococcus carnosus subsp. carnosus* G). *Staphylococcus chromogenes* H). *Staphylococcus cohnii*	59
Figure 2B	GC3 *vs* Enc plot, with each data point representing an individual gene I). *Staphylococcus condimentistr. 58113* J). *Staphylococcus devriesei* K). *Staphylococcus edaphicus str. CCM 8730* L). *Staphylococcus equorum* M). *Staphylococcus felis M23864:W1* N). *Staphylococcus felis* O). *Staphylococcus fleurettii* P). *Staphylococcus gallinarum*	60
Figure 2C	GC3 *vs* Enc plot, with each data point representing an individual gene Q). *Staphylococcus haemolyticus* R). *Staphylococcus hominis* S). *Staphylococcus hyicus* T). *Staphylococcus intermedius NCTC 11048* U). *Staphylococcus kloosii* V). *Staphylococcus lentus* W). *Staphylococcus lugdunensis* X). *Staphylococcus lutrae*	61
Figure 2D	GC3 *vs* Enc plot, with each data point representing an individual gene Y). *Staphylococcus massiliensis S46* Z). *Staphylococcus microti* AA). *Staphylococcus muscae* AB). *Staphylococcus nepalensis* AC). *Staphylococcus pasteuri* AD). *Staphylococcus petrasii subsp. croceilyticus* AE). *Staphylococcus pettenkoferi* AF). *Staphylococcus piscifermentans*	62
Figure 2E	GC3 *vs* Enc plot, with each data point representing an individual gene AG). *Staphylococcus pseudintermedius* AH). *Staphylococcus pseudoxylosus str. S04009* AI). *Staphylococcus rostri* AJ).	63

	Staphylococcus saccharolyticus AK). *Staphylococcus saprophyticus* AL). *Staphylococcus schleiferi* AM). *Staphylococcus sciuri* AN). *Staphylococcus simiae CCM 7213 = CCUG 51256*	
Figure 2F	GC3 *vs* Enc plot, with each data point representing an individual gene AO). *Staphylococcus simulans* AP). *Staphylococcus succinus* AQ). *Staphylococcus vitulinus* AR). *Staphylococcus warneri* AS). Staphylococcus xylosus AT) *Staphylococcus aureus* AU) *Staphylococcus agnetis* AV) *Staphylococcus epidermidis*	64
Figure 3A	GC3 v/s GC12 Plot with each data point representing an individual gene A). *Staphylococcus argenteus str. 58113* B). *Staphylococcus arlettae* C). *Staphylococcus auricularis* D). *Staphylococcus capitis* E). *Staphylococcus caprae M23864:W1* F). *Staphylococcus carnosus subsp. carnosus* G). *Staphylococcus chromogenes* H). *Staphylococcus cohnii*	67
Figure 3B	GC3 v/s GC12 Plot with each data point representing an individual gene I). *Staphylococcus condimentistr. 58113* J). *Staphylococcus devriesei* K). *Staphylococcus edaphicus str. CCM 8730* L). *Staphylococcus equorum* M). *Staphylococcus felis M23864:W1* N). *Staphylococcus felis* O). *Staphylococcus fleurettii* P). *Staphylococcus gallinarum*	68
Figure 3C	GC3 v/s GC12 Plot with each data point representing an individual gene Q). *Staphylococcus haemolyticus* R). *Staphylococcus hominis* S). *Staphylococcus hyicus* T). *Staphylococcus intermedius NCTC 11048* U). *Staphylococcus kloosii* V). *Staphylococcus lentus* W). *Staphylococcus lugdunensis* X). *Staphylococcus lutrae*	69
Figure 3D	GC3 v/s GC12 Plot with each data point representing an individual gene Y). *Staphylococcus massiliensis S46* Z). *Staphylococcus microti* AA). *Staphylococcus muscae* AB). *Staphylococcus nepalensis* AC). *Staphylococcus pasteuri* AD). *Staphylococcus petrasii subsp. croceilyticus* AE). *Staphylococcus pettenkoferi* AF). *Staphylococcus piscifermentans*	70

Figure 3E	GC3 v/s GC12 Plot with each data point representing an individual gene AG). *Staphylococcus pseudintermedius* AH). *Staphylococcus pseudoxylosus str. S04009* AI). *Staphylococcus rostri* AJ). *Staphylococcus saccharolyticus* AK). *Staphylococcus saprophyticus* AL). *Staphylococcus schleiferi* AM). *Staphylococcus sciuri* AN). *Staphylococcus simiae CCM 7213 = CCUG 51256*	71
Figure 3F	GC3 v/s GC12 Plot with each data point representing an individual gene AO). *Staphylococcus simulans* AP). *Staphylococcus succinus* AQ). *Staphylococcus vitulinus* AR). *Staphylococcus warneri* AS). Staphylococcus xylosus AT) *Staphylococcus aureus* AU) *Staphylococcus agnetis* AV) *Staphylococcus epidermidis*	72
Figure 4A	G3s/ G3s + C3s Vs A3s/ A3s + T3s plot with each data point representing an individual gene A). *Staphylococcus argenteus str. 58113* B). *Staphylococcus arlettae* C). *Staphylococcus auricularis* D). *Staphylococcus capitis* E). *Staphylococcus caprae M23864:W1* F). *Staphylococcus carnosus subsp. carnosus* G). *Staphylococcus chromogenes* H). *Staphylococcus cohnii*	75
Figure 4B	G3s/ G3s + C3s Vs A3s/ A3s + T3s plot with each data point representing an individual gene I). *Staphylococcus condimentistr. 58113* J). *Staphylococcus devriesei* K). *Staphylococcus edaphicus str. CCM 8730* L). *Staphylococcus equorum* M). *Staphylococcus felis M23864:W1* N). *Staphylococcus felis* O). *Staphylococcus fleurettii* P). *Staphylococcus gallinarum*	76
Figure 4C	G3s/ G3s + C3s Vs A3s/ A3s + T3s plot with each data point representing an individual gene Q). *Staphylococcus haemolyticus* R). *Staphylococcus hominis* S). *Staphylococcus hyicus* T). *Staphylococcus intermedius NCTC 11048* U). *Staphylococcus kloosii* V). *Staphylococcus lentus* W). *Staphylococcus lugdunensis* X). *Staphylococcus lutrae*	77
Figure 4D	G3s/ G3s + C3s Vs A3s/ A3s + T3s plot with each data point representing an individual gene Y). *Staphylococcus massiliensis*	78

	S46 Z). *Staphylococcus microti* AA). *Staphylococcus muscae* AB). *Staphylococcus nepalensis* AC). *Staphylococcus pasteuri* AD). *Staphylococcus petrasii subsp. croceilyticus* AE). *Staphylococcus pettenkoferi* AF). *Staphylococcus piscifermentans*	
Figure 4E	G3s/ G3s + C3s Vs A3s/ A3s + T3s plot with each data point representing an individual gene AG). *Staphylococcus pseudintermedius* AH). *Staphylococcus pseudoxylosus str. S04009* AI). *Staphylococcus rostri* AJ). *Staphylococcus saccharolyticus* AK). *Staphylococcus saprophyticus* AL). *Staphylococcus schleiferi* AM). *Staphylococcus sciuri* AN). *Staphylococcus simiae CCM 7213 = CCUG 51256*	79
Figure 4F	G3s/ G3s + C3s Vs A3s/ A3s + T3s plot with each data point representing an individual gene AO). *Staphylococcus simulans* AP). *Staphylococcus succinus* AQ). *Staphylococcus vitulinus* AR). *Staphylococcus warneri* AS). *Staphylococcus xylosus* AT) *Staphylococcus aureus* AU) *Staphylococcus agnetis* AV) *Staphylococcus epidermidis*	80
Figure 5	Dinucleotide frequencies of CoPS, CoNS & CoVS	89
Figure 6	Heatmap of Amino Acids Usage of Coagulase positive *Staphylococcus* (CoPS), Coagulase Negative *Staphylococcus* (CoNS) and Coagulase variable *Staphylococcus* (CoVS) groups. (green indicates high usage of amino acid and red indicated low usage of amino acid).	148
Figure 7	Cumulative mean of CoPS, CoNS and CoVS A). GC12 vs GC3 B). ENc vs GC3s C). ENc vs GC12	159
Figure 8	GC% Comparison of *in silico* and *in vivo* analysis	161

LIST OF TABLES

Table Number	Table Title	Page Number
Table 1a	Characteristics of the coagulase positive species of *Staphylococcus* genomes	53
Table 1b	Characteristics of the co-variables of *Staphylococcus* genomes	53
Table 1c	Characteristics of the coagulase negative species of *Staphylococcus* genomes	54-57
Table 2a	Relative dinucleotide abundance analysis for CoPS	82-83
Table 2b	Relative dinucleotide abundance analysis for CoVS	83
Table 2c	Relative dinucleotide abundance analysis for CoNS	83-88
Table 3a	Species-Specific Codon and Codon Pair Usage in Coagulase Positive species	90-91
Table 3b	Species-Specific Codon and Codon Pair Usage in Coagulase -variables	91
Table 3c	Species-Specific Codon and Codon Pair Usage in Coagulase negative species	91-95
Table 4a	Codon usage data of *Staphylococcus agnetis*	98
Table 4b	Codon usage data of *Staphylococcus argenteus*	99
Table 4c	Codon usage data of *Staphylococcus arlettae*	100
Table 4d	Codon usage data of *Staphylococcus aureus*	101
Table 4e	Codon usage data of *Staphylococcus auricularis*	102
Table 4f	Codon usage data of *Staphylococcus capiti*	103
Table 4g	Codon usage data of *Staphylococcus caprae*	104

Table 4h	Codon usage data of *Staphylococcus carnosus*	105
Table 4i	Codon usage data of *Staphylococcus chromogenes*	106
Table 4j	Codon usage data of *Staphylococcus cohnii*	107
Table 4k	Codon usage data of *Staphylococcus condimenti*	108
Table 4l	Codon usage data of *Staphylococcus delphini*	109
Table 4m	Codon usage data of *Staphylococcus devriesei*	110
Table 4n	Codon usage data of *Staphylococcus edaphicus*	112
Table 4o	Codon usage data of *Staphylococcus epidermidis*	113
Table 4p	Codon usage data of *Staphylococcus equorum*	114
Table 4q	Codon usage data of *Staphylococcus felis*	115
Table 4r	Codon usage data of *Staphylococcus fleurettii*	116
Table 4s	Codon usage data of *Staphylococcus gallinarum*	117
Table 4t	Codon usage data of *Staphylococcus haemolyticus*	118
Table 4u	Codon usage data of *Staphylococcus hominis*	119
Table 4v	Codon usage data of *Staphylococcus hyicus*	120
Table 4w	Codon usage data of *Staphylococcus intermedius*	121
Table 4x	Codon usage data of *Staphylococcus kloosii*	122
Table 4y	Codon usage data of *Staphylococcus lentus*	123
Table 4z	Codon usage data of *Staphylococcus lugdunensis*	124
Table 4aa	Codon usage data of *Staphylococcus lutrae*	125
Table 4ab	Codon usage data of *Staphylococcus massiliensi*	126
Table 4ac	Codon usage data of *Staphylococcus microti*	127

Table 4ad	Codon usage data of *Staphylococcus muscae*	128
Table 4ae	Codon usage data of *Staphylococcus nepalensis*	129
Table 4af	Codon usage data of *Staphylococcus pasteuri*	130
Table 4ag	Codon usage data of *Staphylococcus petrasii*	131
Table 4ah	Codon usage data of *Staphylococcus pettenkoferi*	132
Table 4ai	Codon usage data of *Staphylococcus piscifermentans*	133
Table 4aj	Codon usage data of *Staphylococcus pseudintermedius*	134
Table 4ak	Codon usage data of *Staphylococcus pseudoxylosus*	135
Table 4al	Codon usage data of *Staphylococcus rostri*	136
Table 4am	Codon usage data of *Staphylococcus saccharolyticus*	137
Table 4an	Codon usage data of *Staphylococcus saprophyticus*	138
Table 4ao	Codon usage data of *Staphylococcus schleiferi*	139
Table 4ap	Codon usage data of *Staphylococcus sciuri*	140
Table 4aq	Codon usage data of *Staphylococcus simiae*	141
Table 4ar	Codon usage data of *Staphylococcus simulans*	142
Table 4as	Codon usage data of *Staphylococcus succinus*	143
Table 4at	Codon usage data of *Staphylococcus vitulinus*	144
Table 4au	Codon usage data of *Staphylococcus warneri*	145
Table 4av	Codon usage data of *Staphylococcus xylosus*	146
Table 5	Hydrophobicity and aromaticity shape amino acid usage in *Staphylococcus* species	150-158

LIST OF ABBREVIATIONS

Abbreviation	Full Form
A/T	Adenine/Thymine
AD	Atopic dermatitis
ANOVA	Analysis of Variance
Aromo	Aromaticity
AU	Adenine-Uracil
CAI	Codon adaptation index
CBI	Codon bias index
CLABSI	Central line-associated bloodstream infection
COA	Correspondence analysis
CoNS	Coagulase Negative *Staphylococcus*
CoPS	Coagulase Positive *Staphylococcus*
CoVS	Coagulase Variable *Staphylococcus*
CpG	5'—C—phosphate—G—3'
CPS	Codon pair score
CUB	Codon Usage bias
DAMBE	Data Analysis for Molecular Biology and Evolution
DFU	Diabetic foot ulcer
DNA	Deoxy Ribonucleic Acid
ENc	Effective number of codons
Ensembl	European Bioinformatics Institute
GC	Guanine-Cytosine
GRAVY	Grand Average Hydropathicity score
hla	Alpha hemolysin
hlg	Gamma hemolysin
HoCoRT	Host Contamination Removal Tool
MERS	Middle East Respiratory Syndrome
mRNA	Messenger ribonucleic acid
MRSA	Methicillin-resistant *Staphylococcus aureus*
NCBI	National Center for Biotechnology Information
NICU	Neonatal intensive care unit
ORF	Open reading frames
P2	Translational selection
PAMPs	Pathogen-associated molecular patterns
PDB	Protein Data Bank
PR	Parity
PRRs	Pattern recognition receptors

PSMα1	Phenol-soluble modulin alpha 1
RAAU	Relative amino acid usage
rRNA	Ribosomal RNA
RSCPU	Relative Synonymous Codon Pair Usage
RSCU	Relative synonymous codon usage
SARS	Severe Acute Respiratory Syndrome
SCUO	Synonymous codon usage order
SIG	*Staphylococcus intermedius* Group
SLS	Sequence-level selection
SLUSH	*S. lugdunensis* synergistic hemolysins
SPA	*Staphylococcal* Protein A
SPSS	Statistical Package for the Social Sciences
tAI	tRNA adaptation index
TLR-9	Toll-like Receptor 9
tRNA	Transfer ribonucleic acid
TSA	Tryptic Soy Agar
UTI	Urinary tract infection
VEEV	Venezuelan equine encephalitis virus
VRSA	Vancomycin-resistant *Staphylococcus aureus*

LISTS OF APPENDIXES

Appendix Number	Title	Page No.
Appendix I	Letter of Candidacy For Ph.D	186
Appendix II	Publication Details	187-188
Appendix III	Allied Publication Details	189
Appendix IV	Conference Attended Details	190

CHAPTER 1

1. Introduction

The genus *Staphylococcus* encompasses a group of bacteria that are of significant interest due to their diverse clinical and environmental relevance [1]. *Staphylococcus* species are opportunistic bacteria that cause a broad range of infections, from minor skin disorders to serious systemic illnesses. They are a key subject of medical and microbiological study because of their adaptability to many environmental circumstances and the rising prevalence of antibiotic-resistant strains [2]. In addition to their potential for pathogenicity, *Staphylococcus* species are vital to food production, biotechnology, and the normal flora of both humans and animals. To effectively diagnose, treat, and prevent staphylococcal infections, it is essential to comprehend the biology, variety, and pathogenic processes of *Staphylococcus*. [3].

1.1. Background

The genus *Staphylococcus* belongs to the family Micrococcaceae and is characterized by irregularly shaped, coccus-form bacteria with a GC content ranging from 30.7% to 36.4% [4]. These facultative anaerobes, measuring between 0.7 and 1.2 µm, usually form clusters that resemble grape bunches, which is reflected in the name '*Staphylococcus*,' originating from the Greek terms 'staphyle' (bunch of grapes) and 'kokkos' (berry). Staphylococcus species are frequently present as commensal organisms on bird's and warm-blooded animal's skin and mucosal surfaces. [5].

The genus *Staphylococcus* comprises nearly 70 species, each varying in pathogenic potential. Members of this genus are Gram-positive, non-spore-forming, catalase-positive bacteria that typically grow in clusters. The species within this genus are non-motile, coccus-shaped, and not strictly dependent on oxygen for growth, with some exceptions such as *Staphylococcus aureus* subsp. *anaerobius* and *Staphylococcus saccharolyticus*. These bacteria are known to form smooth, rounded convex colonies on culture media and produce a variety of proteins, including hemolysins, nucleases, lipases, coagulases, staphylokinases, and enterotoxins, all of which contribute to their pathogenicity [6].

The capacity of *Staphylococci* to coagulate rabbit plasma determines their classification. Species that produce coagulase enzymes are designated as coagulase-

positive *staphylococci* (CoPS), while species lacking this enzyme are categorized as coagulase-negative staphylococci (CoNS). CoNS are typically commensal bacteria found on mammalian skin but are also prominent causes of soft tissue infections, especially in immunocompromised individuals and hospitalized patients. In contrast, CoPS are often associated with hospital-acquired infections. Moreover, *Staphylococcus* species are known to cause food poisoning and zoonotic diseases, making them of significant concern [7, 8].

1.2. Isolation and History

British surgeon Sir Alexander Ogston first isolated *Staphylococcus* bacteria in 1880 from an infected surgical wound. [9]. Ogston named the genus, and subsequent studies revealed that *Staphylococcus* infections could cause blisters in guinea pigs, mice, and other animals. In 1884, the genus was classified into two distinct species: *S. aureus* and *S. albus*. In 1939, Cowan distinguished *S. epidermidis* as a separate species using coagulase testing, and its characteristics were further defined by serological analysis in 1964. [10, 11]. In 1965, the Baird Parker system was introduced for the classification of Staphylococci and Micrococci. According to this system, Gram-positive and catalase-positive cocci belong to the *Staphylococcus* genus, while members of the *Micrococcus* genus can grow and produce acid from glucose even in the absence of oxygen [12].

Kloos and Sheffler developed a classification scheme in 1975 that described the morphological (e.g., appropriate cell wall structure, non-spore-forming, non-motile) and biochemical characteristics (e.g., fermentative activity, novobiocin susceptibility, nitrate reduction, catalase production) of *Staphylococcus* species. This scheme became a standard in clinical microbiology [13].

Researchers have isolated *Staphylococcus* species from the skin glands, and animal and mammal mucous membranes, as these serve as the natural habitat for the bacteria. *Staphylococci* can also be isolated from various sources, including air, soil, water, surgical equipment, and food products, particularly dairy products [14]. These bacteria exhibit rapid growth in the presence of complex nitrogen sources and under aerobic conditions [15]. Under such conditions, *Staphylococci* produce acetoin. To promote the

growth of *Staphylococci*, certain media can be employed, including Vogel-Johnson agar (VJ), lipovitellin salt mannitol agar (LSM), mannitol salt agar (MSA), blood agar, and Staphylococcus Medium No. 110.. These bacteria can tolerate 5%–10% sodium chloride, a characteristic employed in selective media to inhibit the growth of other bacteria [16].

Certain media can be employed, including Vogel-Johnson agar (VJ), lipovitellin salt mannitol agar (LSM), mannitol salt agar (MSA), blood agar, and *Staphylococcus* Medium No. 110. *Staphylococcus* bacteria are responsible for a broad range of diseases, both foodborne and hospital-acquired, making them a focus of significant research interest [17]. In 2001, the genome of the first species in this genus, *S. aureus*, was sequenced. The genome of the most recently discovered species, *Staphylococcus caledonicus*, was sequenced in 2020. The genomes of 70 species within the genus *Staphylococcus* have now been sequenced, all of which are circular but vary in extrachromosomal elements. The genome sizes among different species range from 2.49 Mb to 2.9 Mb [18].

1.3. Taxonomic History

Initially, the *Staphylococcus* and *Micrococcus* species were classified alongside *Stomatococcus* and *Planococcus* genera within the family Micrococcaceae. However, further molecular and phylogenetic analyses revealed that *Staphylococcus* and *Micrococcus* are not analogous, leading to their separation into distinct groups. The genus *Staphylococcus* forms a coherent group based on DNA-ribosomal RNA (rRNA) hybridization and oligonucleotide analysis of 16S rRNA. In 2004, the genus *Staphylococcus* was reclassified into a new family named Staphylococcaceae [19]. While the *Staphylococcus* species share similarities with members of the Micrococcaceae and Dermacoccaceae families, they have distinct characteristics that warranted this taxonomic revision. The Staphylococcaceae family is now associated with the order Bacillales, suborder Micrococcineae. Currently, more than 70 species and subspecies have been identified under the genus *Staphylococcus* [20].

Staphylococcus species' ability to endure in various environmental settings makes them noteworthy. They can tolerate high salt concentrations, withstanding up to 10% NaCl

in their environment, and can flourish in temperatures as low as 7°C and as high as 48°C. Additionally, most staphylococcal bacteria are susceptible to lysostaphin and exhibit fermentative properties. These bacteria can persist on various non-living surfaces, such as scrub suits, lab coats, aluminum foil, polyvinyl chloride surfaces, countertops, bed rails, stethoscopes, and clothing, for durations ranging from one day to 90 days. Accurate and sensitive diagnostic methods are crucial for identifying *Staphylococcus* species in infections [21].

1.4. Pathogenesis

The pathogenicity of *Staphylococcus* species is largely due to their toxin production and rapid adaptation to environmental changes. Some species have developed resistance to antibiotics, a process governed by genes located on extrachromosomal DNA, such as mobile genetic elements. For instance, strains of *Staphylococcus aureus* have become resistant to methicillin, leading to the emergence of methicillin-resistant *S. aureus* (MRSA) [22]. It has been reported that 90% of *S. aureus* strains are resistant to antibiotic treatment, and in some cases, they have also developed resistance to vancomycin, resulting in vancomycin-resistant *S. aureus* (VRSA). While some species of *Staphylococcus* are highly pathogenic, others are either non-pathogenic or opportunistically pathogenic. Highly pathogenic species, such as *S. aureus*, are typically classified as coagulase-positive, whereas non-pathogenic or opportunistically pathogenic species are categorized as coagulase-negative. *Staphylococcus* bacteria are frequently responsible for hospital-acquired infections, infections related to prosthetic devices and catheters, urinary tract infections (especially in females), soft tissue infections, and food poisoning. In contrast, some species have applications in the food industry [23, 24].

1.5. Coagulase-Positive *Staphylococci* (CoPS)

CoPS are a group of bacteria characterized by their ability to produce coagulase, an enzyme that facilitates blood clotting. This enzyme plays a critical role in the pathogenicity of CoPS, as it helps these bacteria evade the host's immune system by promoting clot formation, which can serve as a protective barrier. The ability to produce coagulase is a significant marker for identifying CoPS, and these species are often

associated with a range of infections in humans and animals. CoPS are responsible for conditions such as skin lesions, toxic shock syndrome, food poisoning, bacteremia, endocarditis, osteomyelitis, pneumonia, abscesses, and meningitis. While some CoPS, like *Staphylococcus aureus*, are well known for causing hospital-acquired infections, other species are more commonly found in animals, with zoonotic potential [25, 26].

1.5.1. *Staphylococcus aureus*

The most well-known member of CoPS, *Staphylococcus aureus*, is regarded as a very dangerous pathogen. From minor skin lesions and soft tissue infections to more serious illnesses like toxic shock syndrome, food poisoning, bacteremia, endocarditis, osteomyelitis, pneumonia, abscess development, and meningitis, this bacteria is well known for causing a wide range of infections in people. *S. aureus* is particularly concerning in hospital environments, where it can lead to severe, potentially life-threatening infections, especially in individuals with weakened immune systems. The ability of *S. aureus* to secrete exotoxins is a key factor in its virulence, as these toxins can damage host tissues and facilitate the spread of the infection. The pathogen is also known for its increasing resistance to antibiotics, including methicillin-resistant *Staphylococcus aureus* (MRSA), which complicates treatment options [25, 26].

1.5.2. *Staphylococcus intermedius*

Staphylococcus intermedius was first isolated from canine animals in 1976 and is primarily a zoonotic pathogen. This species is rarely associated with human infections. *S. intermedius* is named for its characteristics, which are intermediate between those of *S. aureus* and *S. epidermidis*. It is part of the *Staphylococcus intermedius* Group (SIG), which also comprises *S. pseudintermedius* and *S. delphini*. Although not a common human pathogen, *S. intermedius* is important in veterinary medicine due to its presence in animals, particularly dogs. It can cause skin infections in both animals and humans but is primarily a concern in the veterinary field. Its zoonotic potential highlights the need for proper hygiene and precautions when handling animals infected with this bacterium [27].

1.5.3. *Staphylococcus pseudintermedius*

Staphylococcus pseudintermedius is a commensal bacterium commonly found in healthy dogs but can become an opportunistic pathogen, particularly in animals and humans with compromised immune systems. It is frequently isolated from clinical canine specimens, particularly from superficial wounds, where it can lead to skin infections. The bacterium's ability to evolve into a pathogenic form is associated with its growing antibiotic resistance, which presents challenges in treatment. *S. pseudintermedius* can cause infections that range from mild skin issues to more severe conditions in both animals and humans. Its increasing resistance to common antibiotics is an area of concern for veterinary medicine, as well as for the potential for zoonotic transmission [28].

1.5.4. *Staphylococcus delphini*

Staphylococcus delphini is another member of the Staphylococcus intermedius Group (SIG) and is primarily found in carnivorous mammals. This species is generally restricted to infections in animals, particularly those involving suppurative skin lesions. While the pathophysiology of these infections is not fully understood, *S. delphini* has been identified as a potential pathogen in the veterinary context. Its ability to cause skin lesions in carnivorous mammals indicates its role as a pathogen in animal health, but it has rarely been associated with human infections [29].

1.5.5. *Staphylococcus argenteus*

Staphylococcus argenteus is notable for its ability to cause invasive infections, food poisoning, and bone infections globally. One unique characteristic of this species is its inability to produce staphyloxanthin, a carotenoid pigment that is typically responsible for the golden color of *S. aureus* colonies. As a result, *S. argenteus* forms white colonies when cultured on agar plates. Despite its lack of pigment, *S. argenteus* is a significant pathogen and has been implicated in severe infections, including bone infections and foodborne illnesses. It is an emerging pathogen with clinical importance, particularly in its capacity to cause invasive infections in humans [30, 31].

1.5.6. *Staphylococcus schleiferi subsp. Coagulans*

Staphylococcus schleiferi subsp. coagulans is an opportunistic pathogen that primarily causes ear and dermal infections in dogs. It produces several enzymes,

including coagulase, lipase, urease, and β-hemolysin, which contribute to its pathogenicity. While *S. schleiferi* is primarily associated with veterinary infections, it rarely infects healthy humans. Its ability to produce a variety of virulence factors makes it a significant concern in veterinary medicine, though its role in human infections is limited. This subspecies is often found in the skin and mucosal areas of animals, where it can cause localized infections [30, 31].

1.5.7. *Staphylococcus agnetis*

Staphylococcus agnetis is a coagulase-variable species that is commonly isolated from cow milk and is known to cause mastitis in bovine animals. This species is also capable of infecting a wide range of hosts, including canines and poultry. In poultry, *S. agnetis* can lead to severe infections such as osteomyelitis and septicemia. The bacterium's ability to affect multiple species highlights its potential as a zoonotic pathogen, though it is primarily associated with veterinary infections. Its role in mastitis in cows is particularly significant due to the economic impact on dairy farming [30, 31].

1.5.8. *Staphylococcus hyicus*

Staphylococcus hyicus is another coagulase-variable species that is known to cause greasy pig disease in piglets. This disease is characterized by extensive skin damage, including bruising, exudation, erosion, dehydration, and abnormal cell growth in the dermis. *S. hyicus* is a significant pathogen in the pig farming industry, as it leads to considerable economic losses due to the impact on piglet health. While *S. hyicus* infections are largely restricted to pigs, the species has potential for zoonotic transmission, making it important for both animal and human health surveillance [32, 33].

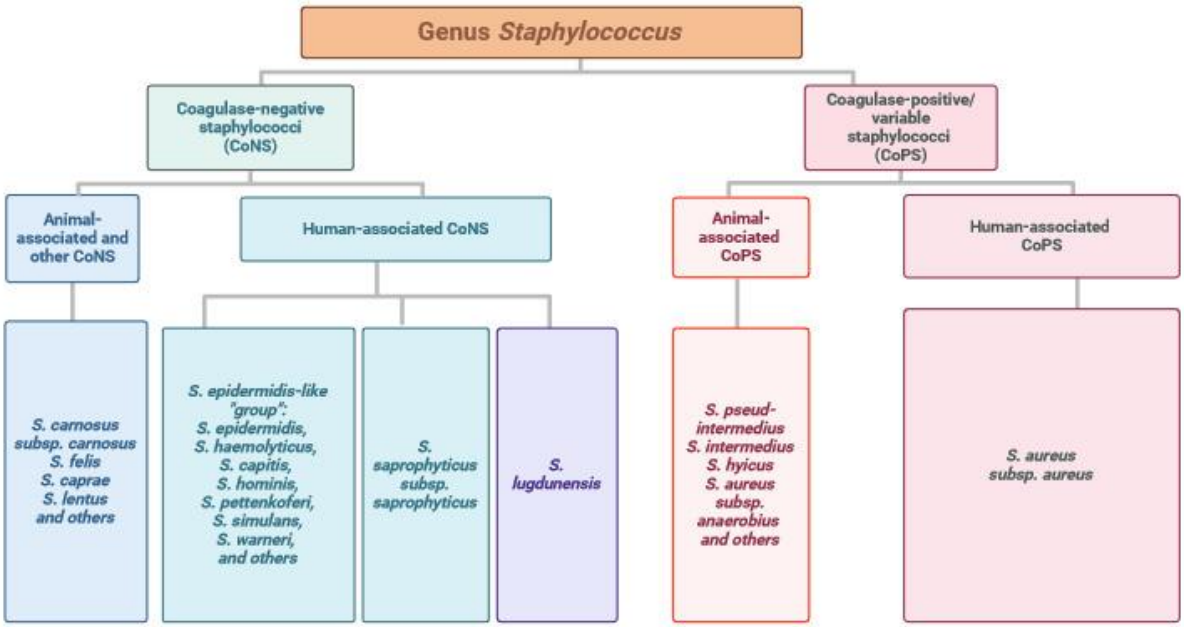


Fig. 1: Classification of Genus *Staphylococcus*

1.6. Coagulase-Negative Staphylococci (CoNS)

CoNS include both pathogenic and non-pathogenic species that do not secrete coagulase. Important members of this group include *Staphylococcus xylosus*, *S. auricularis, S. capitis, S. epidermidis, S. haemolyticus, S. hominis, S. lugdunensis, S. saprophyticus, and S. simulans*. Most CoNS species are either opportunistic mild pathogens or completely apathogenic, though some play key roles in the food industry. Due to their nitrate reductase activity, *Staphylococcus* species contribute to the color and flavor development of food products, particularly in the meat industry. Their antioxidant properties also help prevent rancidity in food [34].

Staphylococcus xylosus is a mutualistic organism found on animal skin and in food. It is a critical starter culture in the food industry, especially in meat processing, but can cause dermatitis in immunocompromised individuals. It can form biofilms and adapt to a wide range of environmental conditions. *Staphylococcus auricularis* is a gram-positive bacillus found in milk, Italian cheese, and as part of the commensal flora of the external auditory canal. Although rarely pathogenic, it can cause endocarditis, anorexia, weight loss, and pleuritic pain. *Staphylococcus capitis* can reproduce under anaerobic conditions and is part of the normal flora on mammalian skin. It is an infrequent cause

of endocarditis and osteomyelitis. *Staphylococcus epidermidis* is also a mutualistic bacterium that resides on human skin [35, 36]. It frequently causes systemic blood infections and infections of sensory organs through contaminated hospital instruments, particularly indwelling catheters. This organism can form a biofilm on the surfaces of artificial devices, a crucial factor in its pathogenicity. Immunocompromised individuals, hospitalized patients, and those who have undergone surgery are at higher risk of acquiring *S. epidermidis* infections. *Staphylococcus haemolyticus* is a commensal bacterium on healthy human skin but can become an opportunistic pathogen. It is the second most common staphylococcal species isolated from clinical samples after *S. epidermidis*. It is a prominent cause of bacterial septicemia and diabetic foot ulcer (DFU) infections, with a higher prevalence in the groin region of hospitalized patients. Its ability to produce biofilms and develop antibiotic resistance makes it a formidable pathogen. *Staphylococcus hominis*, commonly found on the ocular surface, also exhibits antibiotic resistance and biofilm formation. It has been detected on the skin of individuals with atopic dermatitis (AD), and some strains are used in fermentation as starter cultures. *Staphylococcus lugdunensis* is a potential pathogen that affects skin tissues, leading to cutaneous lesions via the release of *S. lugdunensis* synergistic hemolysins (SLUSH). Finally, two coagulase-negative staphylococci, *Staphylococcus saprophyticus* and *Staphylococcus simulans*, are associated with skin and soft tissue infections. *S. saprophyticus* is a common cause of urinary tract infections, particularly in females, while *S. simulans* is a pathogen implicated in bone and joint infections [37, 38].

1.6.1. *Staphylococcus arlettae*

Staphylococcus arlettae is a bacterium frequently found on human skin and mucous membranes, and it's considered a normal part of the body's microbial community. Although it is typically thought to be non-pathogenic, there have been a few isolated cases of it being connected to infections, particularly in people with compromised immune systems or underlying medical disorders. *Staphylococcus arlettae* infections are uncommon, and the pathogenic potential of this bacterium is little understood. When infections do happen, they might present as localized infections or

bloodstream infections, which are frequently linked to medical equipment. Although *Staphylococcus arlettae* is not regarded as a particularly virulent strain, its pathogenicity highlights the necessity of careful infection control procedures and appropriate medical therapy, particularly in healthcare settings and for people who may be more vulnerable to such infections [39, 40].

1.6.2. *Staphylococcus aureus*

Staphylococcus aureus is a bacterium with significant potential to cause various infections in humans. A collection of harmful elements, such as toxins, enzymes, and adhesion proteins, are believed to contribute to its ability to cause disease.

Staphylococcus aureus can lead to a range of infections, from minor skin and soft tissue issues like boils and cellulitis to more severe and possibly life-threatening conditions such as bloodstream infections, pneumonia, and endocarditis. It is also one of the main causes of infections at surgical sites and infections brought on by devices, such as catheter-associated infections. One concerning trait of *S. aureus* is its tendency to develop resistance to drugs, particularly the well-known methicillin-resistant *Staphylococcus aureus* (MRSA). Due to this resistance, therapy becomes more challenging and cautious antibiotic selection is required [41, 42].

1.6.3. *Staphylococcus aureus subsp. anaerobius*

A subspecies of *Staphylococcus aureus*, *Staphylococcus aureus subsp. anaerobius* primarily colonizes the epidermis and nasal passages of ruminant animals. Although it is normally considered non-pathogenic for humans, it can display opportunistic pathogenicity under certain conditions. Infections with *S. aureus subsp. anaerobius* in humans are highly unusual and typically result from contact with infected animals or their substances. Skin and soft tissue infections in a specific area may arise from these conditions, but such instances are uncommon and sporadic [43].

1.6.4. *Staphylococcus capitis*

Usually a component of the skin's normal microbiota, *Staphylococcus capitis* can demonstrate pathogenicity, particularly in medical situations. Its capacity to create

virulence factors, cling to surfaces, and form biofilms is related to its pathogenic potential. *Staphylococcus capitis* is well-recognized for causing infections in vulnerable populations, particularly in neonatal intensive care units (NICUs). Infants born prematurely and those with immature immune systems are particularly vulnerable. Infections, often associated with medical devices, can lead to bloodstream infections and central line-associated bloodstream infections (CLABSIs). Antibiotic resistance makes treating these infections difficult. Strict infection control procedures, focusing on hand hygiene, aseptic techniques, and device-related protocols, are essential in preventing *Staphylococcus capitis* infections in healthcare settings [44, 45].

1.6.5. *Staphylococcus caprae*

Staphylococcus caprae, typically found in the natural bacteria of human skin and mucous membranes, can pose a risk, particularly for individuals with weakened immune systems or existing medical conditions. Its capacity to create virulence factors, cling to surfaces, and form biofilms gives it the potential to be harmful. *Staphylococcus caprae* has been linked to infections in healthcare settings, such as bloodstream infections, urinary tract infections, and surgical site infections, especially after cardiac procedures. It may also result in infections linked to medical equipment that have been implanted [46].

Preventing and treating *Staphylococcus caprae* infections requires stringent infection control procedures, wise antibiotic use, and continuous patient monitoring. Understanding its pathogenicity is essential to reducing the risk of these infections in healthcare settings[47].

1.6.6. *Staphylococcus carnosus*

Staphylococcus carnosus is frequently utilized in the food sector because of its role in the fermentation of meat, where it is primarily recognized as non-pathogenic. The organism lacks the severe virulence factors or processes needed to cause serious illnesses in humans. It thrives in meat and meat products, where it helps in preserving flavor and producing new flavors [48].

1.6.7. *Staphylococcus cohnii*

The bacteria *Staphylococcus cohnii* is frequently found on human mucous membranes and skin. Although it is typically thought of as a normal component of the microbiota, it has the potential to be pathogenic, especially in medical facilities or in people with weakened immune systems. Its capacity to create virulence factors, cling to surfaces, and form biofilms is related to its pathogenic potential. *Staphylococcus cohnii* has been connected to infections in hospital settings, including bloodstream infections and infections involving medical devices. Due to its propensity to acquire antibiotic resistance, these infections can be challenging to treat [49, 50].

1.6.8. *Staphylococcus delphini*

A bacterium called *Staphylococcus delphini* is frequently discovered in dolphins and other marine mammals. Although it is not frequently linked to infections in people, there have been isolated instances of it being involved in opportunistic infections, particularly in people with compromised immune systems or underlying medical disorders. Infections caused by *Staphylococcus delphini* are extremely uncommon, and little is known about the pathogenic potential of this bacterium in humans. Infections can appear locally as infections of the skin or soft tissues or, in rare instances, as bloodstream infections [51].

1.6.9. *Staphylococcus epidermidis*

Normally a benign skin bacterium, *Staphylococcus epidermidis* can turn pathogenic under specific circumstances. It is a major contributor to illnesses connected to healthcare because of its capacity to develop tenacious biofilms on medical equipment. These biofilms offer defense against both antibiotics and the immune system. *S. epidermidis* uses immune evasion techniques, making it difficult for the host to fight it off. Treatment is complicated by antibiotic resistance, particularly in methicillin-resistant forms. It is regarded as an opportunistic pathogen that thrives in the presence of implanted devices or when the host's defenses are weak. Bloodstream infections, endocarditis, surgical site infections, and device-related infections are typical. Effective management involves removing the device (if relevant), using appropriate antibiotics, and adhering to stringent infection control procedures [52].

1.6.10. *Staphylococcus equorum*

A bacterium known as *Staphylococcus equorum* is frequently found in a variety of habitats, including the skin of humans, fermented foods, and horses. It often does not cause infections in healthy people and is widely regarded as non-pathogenic to humans. Although *Staphylococcus equorum* is not frequently linked to human infections, there have been isolated cases of it in people with compromised immune systems or other medical conditions. These infections may manifest as localized infections or bloodstream infections [53].

1.6.11. *Staphylococcus gallinarum*

The bacteria *Staphylococcus gallinarum* is frequently connected to birds and poultry; however, it is not typically regarded as a major human pathogen. It has, however, infrequently been described as an opportunistic pathogen in humans, particularly in those with underlying medical conditions or weakened immune systems. Compared to other Staphylococcus species, *Staphylococcus gallinarum*'s pathogenicity in humans is less well known. Rare infections caused by this bacterium frequently involve bloodstream infections, infections at the site of surgery, or infections involving medical equipment [54].

Even though *Staphylococcus gallinarum* infections are uncommon, they serve as a reminder of the value of upholding good hygiene standards when handling poultry and being watchful around people who are immunocompromised or have implanted medical devices [55].

1.6.12. *Staphylococcus haemolyticus*

A bacterium with emerging pathogenic potential, particularly in healthcare settings, is *Staphylococcus haemolyticus*. Its capacity to create biofilms, stick to surfaces, and acquire antibiotic resistance are all indicators of its pathogenicity. This bacterium is increasingly understood to be a contributor to healthcare-related infections, including bloodstream infections, surgical site infections, and infections linked to medical equipment [56].

Staphylococcus haemolyticus is well known for being resistant to several antibiotics, which makes treatment more challenging. People with compromised immune systems,

those who have had surgery, or those who have had medical devices inserted are frequently affected. To effectively reduce the pathogenicity of *Staphylococcus haemolyticus* infections in hospital settings, strict infection control procedures, prudent antibiotic usage, and monitoring of antibiotic resistance patterns are required [57].

1.6.13. *Staphylococcus hominis*

Staphylococcus hominis is a bacterium frequently found on human mucous membranes and skin. Although it is typically considered a normal component of the microbiota, it can show pathogenicity in some circumstances. Its capacity to create virulence factors, cling to surfaces, and form biofilms is related to its pathogenic potential. *Staphylococcus hominis* has been connected to healthcare-associated infections, notably in people with weakened immune systems or those who have medical implants. These infections may present as bloodstream infections, urinary tract infections, or infections resulting from surgical treatments [58].

In healthcare settings, understanding the pathogenicity of *Staphylococcus hominis* is essential to effectively prevent and manage these infections. This requires stringent infection control procedures, proper antibiotic use, and vigilant monitoring of high-risk patient populations [59].

1.6.14. *Staphylococcus intermedius*

The bacterium *Staphylococcus intermedius* is frequently found on the skin and mucous membranes of animals, especially dogs. Although it is usually not pathogenic to people, it has the potential to cause opportunistic infections, especially in people who are immunocompromised. Rarely, humans may contract infections from *Staphylococcus intermedius*, which can be spread by animals. Bloodstream infections and localized skin and soft tissue infections are among these infections. For people with compromised immune systems who are frequently in contact with animals, understanding its pathogenicity and the risk of zoonotic transmission is crucial [60].

1.6.15. *Staphylococcus kloosii*

Human skin and mucous membranes are frequently colonized by the bacterium *Staphylococcus kloosii*. Although it is typically regarded as a normal component of the

microbiota, it has been linked to sporadic opportunistic infections, particularly in people with compromised immune systems or underlying medical disorders. Its capacity to create virulence factors, cling to surfaces, and form biofilms is related to its pathogenic potential. Infections with *Staphylococcus kloosii* are uncommon, but when they do happen, they can include localized infections or bloodstream infections [61].

1.6.16. *Staphylococcus lentus*

The bacterium *Staphylococcus lentus* is frequently found on human skin and mucous membranes. It is typically considered non-pathogenic; however, it has occasionally been linked to opportunistic infections in people who are immunocompromised or have other underlying medical issues. When infections do happen, they could manifest as localized infections or bloodstream infections. Although the pathogenicity of *Staphylococcus lentus* is not well understood, infections caused by this bacterium serve as a reminder of the value of infection control in populations that are more vulnerable [62].

1.6.17. *Staphylococcus saprophyticus*

A bacterium called *Staphylococcus saprophyticus* is often found in the urogenital tract and is a known culprit in urinary tract infections (UTIs), especially in young women. In comparison to other species, it is less virulent; however, it possesses adhesins that enable it to attach to the urinary system, increasing the risk of infection. Common UTI symptoms include frequent urination, discomfort, and pain brought on by *Staphylococcus saprophyticus*. Since antibiotic resistance is on the rise, therapy primarily consists of antibiotics [63].

1.6.18. *Staphylococcus sciuri*

Although it is typically considered non-pathogenic and is present in animals and the environment, the bacterium *Staphylococcus sciuri* has the potential to be pathogenic under certain circumstances. People with compromised immune systems or underlying health issues are most at risk for opportunistic infections. Human infections are uncommon, but they have been linked to wounds, medical equipment, and bloodstream

infections. The precise pathogenic mechanisms of *S. sciuri* in people are still under investigation [64].

1.6.19. *Staphylococcus simulans*

Staphylococcus simulans is a bacterium that is closely related to *S. aureus* and *S. epidermidis*, but it has lower pathogenic potential. Although it is not frequently regarded as highly virulent, it does produce toxins and enzymes that can harm host tissues and avoid detection by the immune system. Abscesses, wound infections, and, in rare instances, bloodstream infections can result from *Staphylococcus simulans* infections. Patients who have compromised immune systems or have open wounds are particularly at risk for infections. In addition to antibiotics, wound care is essential for the successful treatment of infections [65].

1.6.20. *Staphylococcus vitulus*

The bacterium *Staphylococcus vitulus* is frequently found in cattle and other animals and is typically regarded as non-pathogenic to people. However, in rare cases, it has been linked to infections in people who are immunocompromised or have underlying medical issues. Localized or bloodstream infections are possible manifestations of infections caused by *Staphylococcus vitulus*. However, there is little known about this bacterium's pathogenic potential in humans [66].

1.6.21. *Staphylococcus warneri*

The bacterium *Staphylococcus warneri* is typically found on human mucous membranes and skin. Although it is typically regarded as non-pathogenic, it can cause infections, particularly in people who are immunocompromised or have underlying medical issues.

The ability of *Staphylococcus warneri* to form biofilms has been connected to infrequent instances of endocarditis, bloodstream infections, and infections at surgical sites. Effective infection control and targeted antibiotic therapy are necessary for treating these infections [67].

1.6.22. *Staphylococcus xylosus*

Staphylococcus xylosus is a bacterium frequently found on the skin and mucous membranes of both humans and animals. Additionally, it is used extensively in the food business, particularly in the fermentation of meat. It is not typically linked to human infections. The virulence factors present in more pathogenic *Staphylococcus* species are absent in *Staphylococcus xylosus*. It contributes to the development of flavor and preservation in meat products [68].

1.7. Codon Usage Bias

Codon usage bias (CUB) and amino acid usage analysis are significant in understanding the genetic and evolutionary mechanisms within the *Staphylococcus* genus. Codon usage bias is defined as the non-random usage of synonymous codons to encode the same amino acid. This bias plays a crucial role in gene expression, translation efficiency, and adaptability to environmental stress, ultimately influencing bacterial survival and pathogenicity. The following sections outline the significance, contributing factors, and rationale behind studying CUB and amino acid usage in *Staphylococcus* species [69, 70].

1.7.1. Codon Usage in *Staphylococcus*: Overview

Staphylococcus species exhibit codon usage patterns that are reflective of their genomic composition, primarily influenced by their GC content, evolutionary pressures, and environmental adaptations. Codon usage bias arises from a combination of mutational biases, selection for translational efficiency, and genetic drift. In *Staphylococcus*, these biases are crucial for optimizing gene expression, particularly in genes linked to virulence and survival in hostile conditions such as host immune responses. Factors such as GC content are significant in shaping the codon usage of *Staphylococcus*. Typically, codons with higher GC content are more prevalent in genomes rich in GC, aligning with mutational pressures that favor stability and translational accuracy. This is essential for highly expressed genes, ensuring their optimal translation through codon-tRNA pairing that matches the available tRNA pool. Such pairing minimizes translation errors and enhances efficiency, crucial for maintaining the bacterial load during infections [71, 72].

1.7.2. Influences on Codon Usage Bias

The codon usage within *Staphylococcus* is determined by:

- **GC Content**: Mutational patterns and genomic composition influence the selection of codons. In *Staphylococcus*, codons ending in GC are more frequent in genes that require robust expression.
- **tRNA Abundance**: Codons that correspond to abundant tRNAs are preferred, which supports efficient and error-free protein synthesis. This relationship is often measured using the tRNA adaptation index (tAI) [73].
- **Gene Function and Expression**: High-expression genes tend to utilize codons that correspond to abundant tRNAs, promoting translation efficiency. Genes involved in core functions like metabolism often display a strong codon bias.
- **Selective Pressures**: Evolutionary forces select for codons that not only enhance translational speed but also promote proper protein folding, a factor that is significant in maintaining the functionality of virulence factors [74].

1.8. Amino Acid Usage Analysis

Amino acid usage is intricately linked to the codon usage patterns and reflects the protein requirements of the organism. *Staphylococcus* species may show preferences for certain amino acids due to selective pressures that relate to protein stability, structural requirements, and environmental adaptability. For example, amino acids with simpler biosynthetic pathways or those that contribute to the stability of proteins under stress conditions might be more commonly used. The choice of amino acids can also be a response to the metabolic constraints imposed by the host environment during infection [69, 75].

1.9. Mechanisms Behind Codon Usage Bias

- **Mutation and Selection**: The mutational model suggests that codon usage bias results from random point mutations within the genome that are subsequently subjected to selection. Selection may act on synonymous mutations when they impact translation efficiency or protein function, a process evident in highly expressed genes within *Staphylococcus* [76].

- **Translational Selection**: This theory posits that certain codons are favored to align with abundant tRNAs, reducing the time taken for the ribosome to recognize and bind tRNA. This bias improves the translation elongation rate and accuracy, which is particularly beneficial for pathogenic bacteria like *Staphylococcus* that need to swiftly produce proteins to adapt to host defenses [77].
- **GC Content and Genomic Adaptations**: The overall composition of GC bases influences codon preferences. Higher GC content correlates with stability and robustness in translation, an adaptive feature for pathogens to sustain under stress [78].

1.10. Importance of Codon and Amino Acid Analysis in *Staphylococcus*

Understanding codon and amino acid usage in *Staphylococcus* provides insights into:

- **Pathogenicity and Virulence**: Codon bias may contribute to the differential expression of virulence genes, optimizing their translation during infection and aiding in immune evasion.
- **Antimicrobial Resistance**: Analysis of codon usage could reveal adaptations in genes related to antibiotic resistance, aiding in developing targeted treatments [79].
- **Evolutionary Insights**: Codon usage patterns can serve as markers for evolutionary changes within *Staphylococcus* strains, highlighting horizontal gene transfer or evolutionary pressure from host-pathogen interactions.
- **Biotechnological Applications**: Codon optimization, based on natural biases, can be used to enhance heterologous gene expression in engineered *Staphylococcus* strains for various applications [80].

1.11. Rationale for Studying Codon and Amino Acid Usage in *Staphylococcus*

The study of CUB and amino acid usage in *Staphylococcus* is critical due to the following reasons:

- **Understanding Host-Pathogen Interactions**: Codon bias studies reveal how *Staphylococcus* adapts its gene expression profile in response to host environments. This can help identify potential targets for therapeutic intervention [81].
- **Genetic Engineering**: By comprehending the natural biases, synthetic biologists can design codon-optimized genes that improve the expression of beneficial proteins in microbial systems.
- **Comparative Genomic Studies**: Comparing codon usage across different *Staphylococcus* species or strains can shed light on their evolutionary strategies and adaptability to various niches [82].

Codon and amino acid usage analyses in *Staphylococcus* reveal a complex interplay between mutational pressures, selective forces, and translational mechanisms. These insights not only improve our understanding of the bacterium's genetic makeup and pathogenic strategies but also open avenues for enhanced therapeutic and biotechnological approaches. Staphylococcus species, notably *Staphylococcus aureus*, harbor a group of genes essential to their pathogenicity, facilitating evasion of host immune responses, and virulence [83].

Among these, the *Staphylococcal* Protein A (spa) gene encodes a pivotal protein crucial for immune evasion by specifically binding to the Fc region of immunoglobulins, thus impeding opsonization and subsequent phagocytosis [84]. Additionally, the Phenol-soluble modulin alpha 1 (PSMα1) peptide gene in Staphylococcus aureus enhances biofilm formation and exhibits cytotoxic activity against host cells, aiding in immune evasion and tissue invasion.

It also triggers inflammation and facilitates bacterial dissemination within the host. The Alpha Hemolysin (hla) and Gamma Hemolysin (hlg) genes encode cytolytic toxins capable of disrupting host cell membranes, resulting in tissue damage [85]. Understanding the intricate roles and regulatory mechanisms governing these genes is paramount for devising effective strategies to mitigate Staphylococcus infections, thus underscoring their significance in clinical and public health contexts [84].

An organism's genome contains codons, which are sequences of three nucleotides that match to one of the twenty amino acids, which are the building blocks of proteins. Multiple codons can encode the same amino acid because of the genetic code's inherent redundancy [86]. It is said that these codons are synonymous. However, there is variation in the way these synonymous codons are used, which leads to a phenomena called codon bias. Codon bias is caused by a number of reasons, such as mutational pressure, natural selection, and nucleotide bias. [70, 87].

This bias in codon usage is further associated with various biological phenomena, including gene length, protein amino acid composition, and GC content [87]. This study will examine the specifics of codon usage bias within Staphylococcal species in order to clarify their patterns of usage and any possible consequences for gene expression and evolutionary dynamics. [88].

To elucidate codon usage bias within Staphylococcal species, we employed a comprehensive suite of computational tools and statistical analyses [89, 90]. We have investigated a variety of measures, such as the effective number of codons (ENc), codon bias index (CBI), relative synonymous codon usage (RSCU), and codon adaptation index (CAI), using the CodonW and DAMBE tools. Additionally, we used parity plots, neutrality plots, and correspondence analysis (COA) to fully comprehend codon usage patterns.

Analysing genomic features such as codon usage, genomic composition, and amino acid preferences is vital for deciphering the evolutionary strategies of Staphylococcus bacteria. The preference for specific codons in pathogenic *Staphylococcus* strains can suggests an adaptive mechanism for utilizing host nutrients. This knowledge can give a new insight in the development of targeted antimicrobials that disrupt pathogen-specific metabolic pathways, potentially impeding their ability to thrive in the host environment.

Additionally, knowing how codon use and gene expression relate to one another helps with vaccine and treatment development. Targeting highly expressed virulence genes can disrupt critical pathogenic pathways, thereby reducing the severity and frequency of staphylococcal infections.

Detailed studies on codon and amino acid usage in *Staphylococcus* can provide a roadmap for exploring genetic adaptability and resilience, supporting the development of targeted interventions against staphylococcal infections [91, 92]. Thus, we have designed this work as this research elucidated the molecular mechanisms governing codon usage patterns in *Staphylococcus* species. The present study will explore into the codon usage patterns of 48 distinct *Staphylococcus* species, encompassing those with diverse applications across various industries.

Codon usage bias (CUB) refers to the non-random usage of synonymous codons within the genome, which can vary significantly across different organisms and genes. In bacteria and archaea, CUB is influenced by several factors such as mutation pressures, selective constraints, and environmental adaptations. For instance, studies have shown that acidophilic bacteria exhibit distinct codon preferences linked to their adaptation to extreme environments like copper mines (Hart *et al*., 1997) [97], while pathogenic strains of *Clostridium spp.* tend to have an AT-rich codon usage (Sharma *et al*., 1991) [91]. In plants, CUB is often shaped by factors such as gene stability, GC content, and evolutionary pressures. For example, in the chloroplast genomes of *Panicum* species, AU-rich codons correlate with gene instability (Li *et al*., 1998) [98], while studies in *Cuscuta australis* suggest that pseudogenes exhibit different codon patterns compared to functional genes (Liu X.Y. *et al*., 2001) [101]. Moreover, CUB plays a crucial role in translation efficiency, protein folding, and mRNA regulation, affecting gene expression at the molecular level. In fungi, CUB can impact protein synthesis, with some species showing codon optimization for high-efficiency translation in specific environmental conditions (Baeza *et al*., 2011) [111]. Furthermore, in viral genomes, codon bias is crucial for host adaptation, as seen in the differential codon usage between human and camel hosts in MERS-CoV (Hussain *et al*., 2014) [114]. Overall, CUB provides valuable insights into evolutionary processes, gene functionality, and organismal adaptation, demonstrating its widespread influence across various domains of life and this further discussed in the review of literature.

CHAPTER 2

2. Review of Literature

The study of codon usage patterns across diverse organisms has fascinated researchers due to its intricate connection to evolutionary biology and genetic adaptation. CUB refers to the uneven preference for synonymous codons that code for the same amino acid. This bias is shaped by a complex interaction of factors such as natural selection, mutation pressure, and ecological adaptation. This review of literature aims to delve into the various studies that highlight how CUB can reflect an organism's evolutionary history, genomic efficiency, and environmental adaptability. From the selective pressures shaping codon preferences and the coevolution of tRNA gene content to the implications for translation efficiency, CUB offers a window into the adaptive strategies employed by different species. Understanding these patterns not only provides insights into the genetic and evolutionary dynamics of organisms but also underscores the broader implications for biotechnology, taxonomy, and genomic annotation [93-95]. Here, we have discussed below this collective research that forms a foundation for continued exploration into the evolutionary significance of codon usage across life forms.

2.1. Codon Usage Bias Across Domains of Life

Novoa, E.M. *et al.,* explored the historical patterns of codon usage among various organisms, revealing that synonymous codons were not utilized uniformly. The study indicated that selective pressures played a significant role in shaping codon preferences, while GC content was identified as a primary factor influencing codon usage variation across different species. The authors found that highly expressed genes exhibited a pronounced codon bias, which correlated with the availability of specific tRNAs. Furthermore, the coevolution of tRNA gene content and codon usage was discussed, suggesting that this relationship had implications for translation efficiency in the past. The research also highlighted that codon usage signatures could be employed to classify genomic sequences into their respective domains of life, with a particular focus on arginine codons as key indicators of these distinctions. This understanding of codon usage patterns provided valuable insights into the evolutionary relationships among species and contributed to the field of taxonomic annotation in genomics. Overall, the

findings underscored the complexity of codon usage and its evolutionary significance across different domains of life [96].

Five acidophilic bacteria that were isolated from copper mines and known to be able to solubilize copper were found to contain CUB by Hart, A. *et al.* The researchers aimed to understand how CUB reflected the genomic adaptations of these bacteria to extreme environmental conditions. They found that acidophilic bacteria generally exhibited low CUB, suggesting a random usage of synonymous codons, which was consistent with their capacity to thrive in diverse habitats and their slower growth rates. This finding aligned with previous studies that indicated low CUB could enhance metabolic variability, thereby aiding adaptation to harsh environments characterized by high metal concentrations. Overall, the study helped shed light on the relationship between CUB and environmental adaptation in acidophilic bacteria, highlighting the potential for these microorganisms in biotechnological applications, especially in biomining processes. The analysis also found significant differences between the consortium and non-consortium strains in the unique sets of genes related to metal and oxidative stress resistance, with these differences being especially noticeable in categories associated with metal resistance and iron and sulfur oxidation. The results showed that the lower values of CUB in the consortium's unique genes suggested a higher transcriptional adaptation to extreme conditions, likely acquired as a survival strategy in the metal-rich mining environment [97].

Sharma, A and group conducted a thorough examination of the codon usage patterns among various species within the genus Clostridium. They highlighted the significant diversity of this genus, which included both pathogenic species, such as *C. botulinum* and *C. tetani*, and non-pathogenic species like *C. pasteurianum*, known for its role in nitrogen fixation. This diversity underscored the varying ecological roles these bacteria played, impacting both health and industrial applications. The authors discussed the pathogenic characteristics of several Clostridium species, emphasizing *C. difficile* as a major cause of hospital-acquired infections and *C. perfringens*, which is associated with gas gangrene. They noted the presence of virulence factors and the role of horizontal gene transfer in the evolution of these pathogens, which contributed to their adaptability and survival in different environments. Previous genomic studies had focused on

metabolic pathways and molecular aspects of Clostridium species, but the authors identified a gap in research regarding codon and amino acid usage. Their study aimed to fill this gap by revealing that pathogenic Clostridium species exhibited a distinct CUB, favoring AT-rich codons. This preference was linked to their smaller genome sizes and lower biosynthetic costs, which were crucial for their survival and pathogenicity in host environments. The authors also provided insights into the evolutionary implications of codon usage, suggesting that the pathogenic characteristics of Clostridium species had become defining features. Phylogenetic analyses indicated that species clustered based on their pathogenicity, highlighting a significant evolutionary trend within the genus. In conclusion, the authors proposed that understanding codon usage and its implications for protein synthesis could offer valuable insights into the evolutionary trajectories of Clostridium species, particularly regarding their adaptations to host interactions and environmental challenges. This literature survey emphasized the importance of codon usage analysis in elucidating the genomic and evolutionary dynamics of pathogenic versus non-pathogenic Clostridium species [91].

Arella, D. and co-workers investigated the intricate relationship between CUB and the ecological adaptation of microbial species. It highlighted that each species exhibited unique codon biases, which emerged from evolutionary trade-offs involving translation efficiency, biosynthetic costs, and nutrient availability in their respective environments. An comprehensive collection of 615 microbiological organisms including 71 archaea and 544 bacteria was used in the study. Using principal component analysis, it was shown that species with similar environmental factors and phenotypic traits had similar preferences for codons. The findings indicated that organisms thriving in multiple habitats, such as facultative organisms and mesophiles, demonstrated reduced translational efficiency, as measured by the average tRNA adaptation index (tAI). This suggested that their codon choices were influenced by the need to adapt to diverse environments. The study also noted that the availability of tRNA molecules played a crucial role in shaping CUB, as organisms with higher tRNA gene copy numbers tended to exhibit stronger translational selection. Furthermore, the research underscored the significance of horizontal gene transfer in microbial evolution, particularly in

hyperthermophilic bacteria, which adapted to extreme environments by acquiring genes from other organisms. Overall, the study provided a comprehensive analysis of how codon usage and translational efficiency are linked to the ecological niches occupied by microbial organisms, emphasizing the need for further research to explore the underlying mechanisms of these adaptations [98].

Fu, Y. and research group studied CUB in ciliated protozoa and provided significant insights into the genetic mechanisms and evolutionary history of these organisms. It highlighted that CUB is influenced by various factors, including mutation pressure and natural selection, which shape the codon preferences observed in different species. Previous research had established that ciliates, such as *Tetrahymena thermophila* and *Paramecium tetraurelia*, exhibited distinct patterns of CUB, but a comprehensive analysis across multiple ciliate species was lacking. This study aimed to fill that gap by analyzing 21 ciliate species from various classes and subphyla, revealing that most species preferred codons ending in A or T, contrasting with other organisms like plants that showed a bias for GC-ending codons. The research employed methodologies such as Parity Rule 2 (PR2) plot analysis and Neutrality plot analysis to explore the influences on CUB, demonstrating that mutation pressure significantly affected codon usage, although natural selection appeared to play a more substantial role. The findings indicated that the GC content in the macronuclear genomes of the studied ciliates was generally below 50%, and the base compositions of GC and GC3s were markedly distinct. This study not only contributed to the understanding of CUB in ciliates but also suggested implications for optimizing gene editing techniques in these model organisms. Overall, the research underscored the complexity of factors influencing CUB and provided a foundation for future studies aimed at unraveling the molecular evolution of ciliates and their adaptation mechanisms to environmental changes [99].

Smith, R. explored the phenomenon of CUB, which had been recognized as a significant aspect of molecular biology. It highlighted that CUB referred to the uneven frequency of synonymous codons used in different organisms and genes. The authors noted that two primary mechanisms drove this bias: mutational biases and selection forces. Mutational biases altered codon frequencies, particularly through changes in the

third nucleotide position of codons, leading to preferences for certain codons based on nucleotide composition. Selection forces favored specific synonymous codons due to advantages in translation efficiency, especially in highly expressed genes. The paper emphasized the influence of G+C content on CUB, indicating that codons with higher G+C content were often more prevalent, influenced by processes like GC-biased gene conversion. Various metrics for measuring CUB were discussed, particularly the effective number of codons (N_c). The authors presented new measures, such as N_c (1), N_c (2), and N_c (3), which incorporated G+C content and other evolutionary factors affecting codon usage. The findings suggested that variation in CUB across different organisms was primarily driven by mutational forces, while variation within genomes was more influenced by selectional forces. This distinction was crucial for understanding the evolutionary dynamics of CUB. The paper formalized the understanding of CUB by distinguishing the various forces at play and providing new metrics for analysis. This contributed to a deeper understanding of how codon usage was shaped by both mutational and selection pressures across different biological contexts. Overall, the research provided valuable insights into the complexities of CUB and its implications for evolutionary biology [100].

2.2. Codon Usage Bias in Plants

Li, G and colleagues revealed significant insights into the evolutionary dynamics of these plants. Previous research had often focused on limited aspects of genetic diversity, primarily examining specific genes or a small number of species, which resulted in a fragmented understanding of the overall genetic landscape of Panicum. The present study aimed to fill this gap by analyzing nineteen chloroplast genomes, employing various computational bioinformatics techniques to evaluate genetic diversity and codon usage patterns, including the relative synonymous codon usage (RSCU), codon adaptation index (CAI), and effective number of codons (ENC). The results showed that the majority of the chloroplast genomes were AU-rich. This was in line with other research that demonstrated how natural selection affects codon bias in plant genomes. Furthermore, the research underscored the diversity of specific genes, noting that shorter genes exhibited greater instability, a phenomenon that had not been thoroughly addressed in earlier studies. The study also pointed out that while the overall

evolutionary differences among the chloroplast genomes were not significant, the genetic diversity of typical genes varied considerably. This highlighted the need for further exploration of the factors influencing the evolution of specific genes within the chloroplast genomes of Panicum species. The results contributed to a deeper understanding of the evolutionary pressures acting on these plants and provided a foundation for future research in molecular breeding and conservation efforts. Overall, the study not only advanced the knowledge of codon usage patterns in Panicum but also emphasized the importance of comprehensive analyses in uncovering the complexities of genetic diversity in plant species [101].

An extensive examination of codon use patterns in Cuscuta australis' protein-coding sequences and pseudogenes was conducted by Liua, X.Y., and their study team. According to their findings, high-frequency codons tended to favor A/U in the third position in both types of sequences. The preferred codons, however, showed a substantial difference: pseudogenes preferred A/U-ending codons, while protein-coding sequences tended to employ G/C-ending codons. This discrepancy suggested *that C. australis'*s codon use patterns were significantly influenced by nucleotide composition. The scientists used a number of methods, such as neutrality plots and relative synonymous codon use (RSCU), to investigate the effects of natural selection and mutation pressure. Their results showed that in both protein-coding sequences and pseudogenes, natural selection had a greater impact on codon use patterns than mutation pressure. The study also showed that CUB and gene expression levels were positively correlated, with highly expressed protein-coding genes showing higher codon bias than their weakly expressed counterparts. This correlation was consistent with observations in other species, such as *Arabidopsis thaliana*. The authors also noted that the expression levels of pseudogenes were generally low, attributed to factors such as sequence inactivation and poor transcription efficiency. Overall, this research provided valuable insights into the codon usage patterns in *Cuscuta australis*, emphasizing the evolutionary implications of codon bias in relation to the plant's parasitic lifestyle and adaptation strategies [102].

Zhang, P. and associates showed that natural selection and mutation pressure are two of the many elements that influence CUB. This study filled a major research gap by being

the first to comprehensively analyze codon use patterns within the chloroplast genomes of nine *Gynostemma* species. A predilection for A and T nucleotides, particularly in the third codon position, was shown by the study, which showed that the GC content of these genomes was less than 50%. There may have been a common pattern in the development of chloroplast genes because this A/T bias reflected patterns seen in other plant species. The study confirmed previous findings that highlighted the significance of codon optimization for effective gene expression by identifying 12 ideal codons and 29 high-frequency codons, the majority of which terminated in A or T. The authors used a variety of multivariate statistical techniques, such as correspondence analysis and neutrality plots, to examine the variables affecting codon use. According to their findings, mutation pressure had less of an impact on codon use patterns than natural selection. All things considered, the study added important knowledge on CUB in *Gynostemma* species and served as a benchmark for further research on the genomes of other plant species' chloroplasts. The results provided a better understanding of the molecular history of these commercially significant plants by highlighting the intricacy of codon use patterns and the interaction of evolutionary factors [103].

Chakraborty, S. and colleagues revealed significant insights into the genetic and evolutionary processes affecting Oryza species. CUB, which refers to the non-uniform usage of synonymous codons, was found to be influenced by mutation pressure and natural selection. The research highlighted that while CUB had been extensively studied in nuclear genomes, its implications in organellar genomes, particularly chloroplasts, had not been thoroughly explored. The chloroplast genome, known for its circular structure and essential role in photosynthesis, was analyzed across 18 Oryza species, showcasing the genetic diversity within this genus. In contrast to earlier research on nuclear genes in other plant groups, the results showed that the chloroplast genes were AT-rich, with a significant preference for A- and T-ending codons. Key metrics assessed in this approach were the effective number of codons (ENC) and relative synonymous codon use (RSCU), which showed that various species' levels of mutation and selection pressures influenced CUB. The study also found a strong relationship between nucleotide composition and synonymous codon use order (SCUO), highlighting these species' evolutionary adaptations. According to the

findings, Oryza species' evolutionary history and environmental impacts were represented in the patterns of codon use. All things considered, this study advanced our knowledge of the evolutionary processes of Oryza's chloroplast genes and highlighted the significance of CUB in clarifying the genetic connections and adaptations of these essential agricultural plants [104].

By examining CUB in the genomes of tobacco, tomato, and potato, Anwar, A.M., and his research team were able to uncover important information on the evolutionary processes of these species. It became clear that CUB is not evenly distributed and is affected by a number of variables, like as mutation and natural selection. To evaluate CUB, the study used a range of techniques, such as the translation selection (P2) index, relative synonymous codon use (RSCU), codon adaptation index (CAI), and effective number of codons (ENc). These techniques offered a thorough comprehension of the formation of codon use patterns. Prior research revealed that little was known about CUB in the Solanaceae family, especially in relation to tomatoes. The results suggested that GC3 concentration and CAI were important determinants influencing codon use. Additionally, the research identified 26 optimal codons that predominantly ended with T or A, indicating a preference for these codons across the genomes. This finding aligned with previous studies that reported similar trends in other species, suggesting a broader pattern of codon preference within the Solanaceae family. The literature survey underscored the importance of understanding CUB in the context of evolutionary biology and gene expression, particularly in economically significant crops like tobacco, tomato, and potato. The findings contributed to a growing body of knowledge regarding the genetic and evolutionary dynamics within the Solanaceae family, emphasizing the need for further research to explore the implications of CUB on gene function and adaptation in these important agricultural species. Overall, the study provided valuable insights into the factors influencing CUB and its evolutionary significance [105].

In line with studies in other species like *Medicago truncatula* and *Oryza sativa*, Gao, Y. and colleagues stressed that natural selection and mutation pressure are the primary processes causing CUB, with mutation acting as the dominating driver. The effective number of codons (ENC) and codon adaptation index (CAI) values, which showed a

weak codon bias and low expression levels, further supported the analysis's finding that the WRKY genes had an average GC content of 43.42%, showing a preference for A/T(U) ending codons. The study assessed the impact of mutation and selection on codon use using a number of analytical techniques, such as PR2-plot and ENC-plot analyses. According to the PR2 study, AT and GC were used disproportionately, indicating that mutational pressures were primarily responsible for codon use preferences. The idea that mutational bias contributes to a comparable GC content across all codon positions is further supported by the substantial positive association that was found between GC12 and GC3. The findings aligned with previous research on WRKY genes in other plants, demonstrating varying degrees of CUB influenced by both mutation and selection. This systematic analysis of CUB in H. annuus WRKY genes not only contributed to the understanding of the genetic architecture of this species but also provided a theoretical foundation for future transgenic studies aimed at optimizing codon usage. Overall, the research underscored the complexity of CUB and its implications for plant genetics and breeding programs [106].

2.3. Codon Usage Bias in Fungal Systems

Xu, Y. and co-workers studied CUB in Pichia pastoris and revealed significant insights into its role in protein synthesis. Researchers established that CUB, which refers to the non-random selection of codons in coding sequences, was a critical factor influencing the rate of protein synthesis. They noted that more frequent codons were associated with highly expressed genes, facilitating faster translation due to the abundance of corresponding tRNAs, while rare codons were linked to slower translation rates and were typically found in lowly expressed genes. The research highlighted the unique codon preferences of P. pastoris, particularly its high frequency of A/U in codons, which complicated the effects of codon optimization on protein expression and folding. Additionally, the study examined the impact of codon usage on signal peptides, finding that while the choice of codons in N-terminal signal peptides did not significantly affect their secretion ability, it did influence overall protein expression levels. This indicated that codon optimization strategies should consider the specific characteristics of signal sequences. Furthermore, the research demonstrated that different genes exhibited variable responses to codon optimization, with structural disorder being a crucial factor

in these responses. For some genes, an updated codon optimization approach that accounted for protein structure and transcriptional factors was deemed necessary to ensure optimal expression and conformation. Overall, the findings underscored the importance of codon bias in regulating gene expression and protein conformation in P. pastoris, advocating for a more nuanced understanding of codon optimization to improve protein expression success rates in this yeast system [107].

Wang, F. and colleagues investigated the CUB in mitochondrial protein-coding genes across 12 species of Candida. It highlighted the increasing incidence of fungal infections, particularly those caused by Candida species, which posed significant challenges in clinical treatment due to drug resistance and adverse effects. The researchers focused on the mitochondrial genes, as they play a crucial role in the pathogenicity and drug resistance of fungi. They employed various analytical methods, including RSCU, CAI, and PR2 analysis, to assess the codon usage patterns. The results indicated that the mitochondrial genes predominantly favored A/T bases over G/C bases, with specific codons like UUA, AGU, and CCU being commonly preferred across the species. The study also found a significant correlation between GC content and CUB, suggesting that GC content was a critical factor influencing gene expression levels. Furthermore, the analysis revealed that both selection pressure and mutation pressure played roles in shaping CUB, with mutation pressure being the dominant factor. By analyzing the CUB, the study aimed to provide insights that could guide future antifungal research and therapeutic strategies. Overall, the research contributed to the growing body of knowledge regarding the genetic factors influencing the virulence of Candida species and highlighted the need for continued investigation into the evolution of fungal pathogenic genes. This understanding could potentially lead to novel approaches in reducing pathogenicity by altering codon usage patterns in these organisms [108].

In the study of CUB in yeasts, Baeza, M. and research group analyzed the relationship between CUB and various biological factors, including gene expression levels, growth temperature, and protein structure. They found that CUB was a significant phenomenon across different yeast species, with certain codons being preferentially used over others. This bias was particularly evident in highly expressed open reading frames (ORFs),

which tended to exhibit a higher content of preferred codons, enhancing translation efficiency. The researchers noted that the maximum growth temperature of yeast strains influenced their CUB patterns, with yeasts from colder environments favoring codons that supported survival in low temperatures. This correlation was supported by transcriptomic analyses of Antarctic yeasts, which demonstrated distinct CUB patterns linked to their environmental adaptations. Furthermore, the study revealed that the differences in relative synonymous codon usage (RSCU) values were more pronounced between groups of genes with varying expression levels than between those with different GC content. The implications of CUB extended to protein structure, as it influenced translation accuracy and efficiency, thereby affecting cotranslational protein folding and mRNA stability. A comprehensive analysis involving 89 yeast strains from diverse environments was conducted to evaluate these correlations, aiming to deepen the understanding of how CUB interacts with gene expression, growth temperatures, and protein properties. Overall, the findings underscored the multifaceted nature of CUB in yeasts and its significance in adaptation and protein synthesis processes [109].

2.4. Functional and Mechanistic Insights into Codon Usage Bias

Liu, Y. *et al.,* explored the critical role of CUB in gene expression and protein folding. It established that CUB, which refers to the preference for certain synonymous codons, was not merely a silent phenomenon but had significant implications for cellular processes. The authors highlighted that codon usage influenced translation elongation speed, thereby regulating translation efficiency and accuracy, which were essential for proper protein folding in various biological contexts. They discussed how different synonymous codons could lead to nonuniform ribosome decoding rates on mRNAs, affecting the cotranslational folding process critical for protein functionality. Additionally, the authors highlighted that codon usage played a key role in determining mRNA levels, influencing both translation-dependent mRNA degradation and translation-independent transcriptional and posttranscriptional mechanisms. The authors also examined the regulation of tRNA expression, noting that variations in tRNA levels among different tissues contributed to tissue-specific gene expression. Despite the advancements in understanding codon usage, authors identified several unanswered questions regarding its diverse functions and mechanisms, particularly its

influence on translation initiation and adaptation to different protein folding processes. The authors concluded that while there was substantial evidence supporting the importance of codon usage in gene regulation, further research was necessary to fully elucidate its physiological relevance and potential links to human diseases. Overall, they provided a comprehensive overview of the multifaceted roles of codon usage, underscoring its significance as a layer of genetic information that influences gene expression and protein function across various organisms [110].

Deng, Y. *et al*., recognized that while synonymous codons carry equivalent information, their usage was not random, leading to questions about the underlying causes of this bias. Various heuristic measures, such as the frequency of optimal codons (F opt) and codon adaptation index (CAI), were employed to quantify CUB, but these methods often relied on specific assumptions about the organisms being studied. Some models focused on the effects of mutation bias and selection for translational efficiency, yet they typically considered a limited range of evolutionary forces. The literature distinguished between two main selection types: beanbag selection, which suggested uniform selection across the genome, and sequence-level selection (SLS), which indicated that selection could vary among individual genes. As genomic data became more abundant, the need for new approaches to study CUB emerged, particularly for organisms with limited data. The paper introduced a novel quantification method, the distance measure D, which did not depend on predefined reference sets, allowing for broader applicability across diverse genomic data. The model drew from statistical mechanics and stochastic thermodynamics, providing a comprehensive framework for understanding CUB without relying on specific selection drivers. The analysis of approximately 1500 genomes across three microbial kingdoms revealed that SLS significantly influenced overall CUB, while also uncovering variability in amino acid-specific codon usage patterns across different branches of the tree of life, correlating with organismal tRNA content. This research contributed valuable insights into the complex dynamics of CUB and its evolutionary implications [111].

Tang, D. and co-workers recognized that while synonymous codons carry equivalent information, their usage was not random, leading to questions about the underlying causes of this bias. Various heuristic measures, such as the frequency of optimal codons

(F opt) and codon adaptation index (CAI), were employed to quantify CUB, but these methods often relied on specific assumptions about the organisms being studied. Some models focused on the effects of mutation bias and selection for translational efficiency, yet they typically considered a limited range of evolutionary forces. The literature distinguished between two main selection types: beanbag selection, which suggested uniform selection across the genome, and sequence-level selection (SLS), which indicated that selection could vary among individual genes. As genomic data became more abundant, the need for new approaches to study CUB emerged, particularly for organisms with limited data. The authors introduced a novel quantification method, the distance measure D, which did not depend on predefined reference sets, allowing for broader applicability across diverse genomic data. The model drew from statistical mechanics and stochastic thermodynamics, providing a comprehensive framework for understanding CUB without relying on specific selection drivers. The analysis of approximately 1500 genomes across three microbial kingdoms revealed that SLS significantly influenced overall CUB, while also uncovering variability in amino acid-specific codon usage patterns across different branches of the tree of life, correlating with organismal tRNA content. This research contributed valuable insights into the complex dynamics of CUB and its evolutionary implications [112].

Khandia, R. *et al*., investigated the relationship between CUB and gene length in genes linked to neurodegenerative disorders. It highlighted those neurodegenerative diseases, such as Alzheimer's and Parkinson's, have genetic and environmental risk factors, with aging being a significant contributor. The study emphasized that CUB is a crucial aspect of molecular characterization, influencing gene evolution and expression. It was noted that longer genes typically exhibited a higher codon bias, which was believed to enhance translational efficiency due to selection pressure. The research also explored how gene length correlated with expression levels, revealing that shorter genes were associated with higher expression rates, while longer genes were linked to more complex functions. The nucleotide composition of the genes was analyzed, showing that the G nucleotide was the most abundant, while T was the least prevalent. This compositional bias was significant in understanding how gene length influenced CUB. The study identified specific codons that were positively or negatively associated with

gene length, such as TTA and GTT showing positive correlations, while GTA and AGC exhibited negative correlations. Furthermore, the findings indicated that the association between CUB and gene length varied with segment size, with segments showing a positive correlation. Overall, authors provided a comprehensive overview of the interplay between gene length and CUB, underscoring their importance in understanding gene function and potential therapeutic strategies for neurodegenerative diseases [113].

Zhao, F. and colleagues investigated the genome-wide impact of codon usage on transcription and identified potential regulators affecting gene expressionThey discovered that the preference for specific synonymous codons, known as codon use bias (CUB), was a key factor in controlling gene expression in many species. The scientists found 18 potential components whose deletion decreased the association between RNA levels and codon use through a thorough genetic screen in Neurospora. This suggested that these elements were important mediators of the transcriptional impacts of codon use. Chromatin regulators and transcription factors, including the H3K36 methyltransferase, which is known to affect transcription elongation and chromatin structure, were among the factors found. The study also highlighted that optimal codons correlated positively with mRNA levels, while rare codons exhibited negative correlations, suggesting that codon optimality was essential for determining transcription levels. The researchers utilized nuclear RNA sequencing to eliminate translation-dependent effects, confirming that the observed correlations were indeed transcriptional rather than translational. They found that the correlation coefficients between CUB and mRNA levels were strong, reinforcing the idea that codon usage influences transcription independently of translation. The findings underscored the complexity of transcriptional regulation, suggesting that codon usage might interact with various transcriptional regulatory mechanisms. Overall, the study provided significant insights into how codon usage affects gene expression at the transcriptional level, paving the way for future research into the intricate relationships between codon usage, chromatin dynamics, and transcriptional regulation [114].

Zhao, Y. and co-workers explored the role of CUB in the evolutionary dynamics of vesicular glutamate transporters (VGLUTs), which consist of three isoforms: VGLUT1,

VGLUT2, and VGLUT3. Previous research indicated that synonymous mutations, often considered nonfunctional, could influence gene expression and protein stability, thereby affecting evolutionary rates. The authors noted that while many scholars focused on non-synonymous substitutions, the impact of synonymous mutations on protein evolution was gaining attention. With a negative association found between codon preference and expression in SLC17 family proteins, they emphasized that codon choice might have a substantial impact on gene expression levels. In order to comprehend how variations in codon use patterns contributed to variations in the isoforms' functional relevance, expression levels, and distribution ranges, the study employed bioinformatics techniques to examine the coding sequences of VGLUTs across several species. The findings showed that VGLUTs with higher codon preference had lower levels of expression and functional relevance, indicating that codon use may play a crucial role in their evolutionary paths. The evolutionary links between VGLUTs were also covered by the authors, who pointed out that VGLUT1 and VGLUT2 were more closely related than VGLUT3, which showed notable variations in expression and functional functions. The study's conclusion, which emphasized the need for more investigation to elucidate these interactions, was that codon use preference was a possible factor impacting VGLUT developmental expression and protein evolution. Overall, the findings contributed to a deeper understanding of how synonymous codon usage could shape gene expression and evolutionary processes in the context of VGLUTs, highlighting the importance of considering CUB in evolutionary biology studies [115].

2.5. Codon Usage Bias in Viruses

Kumar, N. *et al.,* examined the evolutionary dynamics of coronaviruses (CoVs) that were hosted by bats, which served as natural reservoirs for many viruses affecting humans. It classified coronaviruses into four genera: Alpha-, Beta-, Gamma-, and Delta coronaviruses, emphasizing that alpha- and beta-CoVs were primarily found in bats, while gamma- and delta-CoVs were more common in birds. The authors highlighted that viruses, including CoVs, were subject to evolutionary pressures from their hosts, influencing the genetic and biochemical adaptations crucial for their survival and transmission. They investigated codon usage patterns in CoVs, noting that these

patterns were influenced by host species and their specific cellular environments. The study employed genetic tools to analyze these biases, focusing on factors like CpG dinucleotide content and relative synonymous codon usage (RSCU). The findings suggested that while bats harbored a diverse range of CoVs, the direct spillover of SARS-CoV-2 from bats to humans was unlikely without an intermediary host. This underscored the importance of understanding the ecological and behavioral factors contributing to zoonotic transmission. The authors called for enhanced surveillance of bat populations to identify viruses with zoonotic potential and to better understand the ecological factors that facilitated cross-species transmission [116].

Wang, H. and colleagues highlighted that Venezuelan equine encephalitis virus (VEEV) exhibited a relatively low CUB, which was assessed through various metrics such as the ENc, CAI, and RSCU. The research also emphasized the adaptation of VEEV to its primary rodent hosts, particularly M. auratus, which demonstrated the highest normalized CAI values, suggesting a co-evolutionary relationship between the virus and its hosts. Furthermore, the study noted that the codon usage patterns of VEEV were similar to those of its long-term hosts, including humans and horses, which underscored the importance of understanding these patterns for controlling cross-species transmission. The implications of this research extended to the development of antiviral strategies and vaccines, as the CUB could inform the design of attenuated virus strains for clinical use. Additionally, the study contributed to the broader understanding of how translational selection influences the evolution of RNA viruses, drawing comparisons with other viruses like ZIKV and SARS-CoV-2. Overall, this research not only advanced the knowledge of VEEV's evolutionary features but also provided valuable insights for future studies aimed at preventing outbreaks and enhancing public health responses to viral infections [117].

Hussain, S. and colleagues conducted a study on the Middle East Respiratory Syndrome Coronavirus (MERS-CoV), offering valuable insights into codon usage and nucleotide bias within its genes. The team analyzed 4,751 MERS-CoV genes to uncover the forces influencing codon usage bias (CUB). They found that CUB was generally low but highly conserved across genes, with natural selection playing a dominant role, while mutation pressure was a lesser influence in certain genes. The study showed that codon

usage was not random; instead, it was shaped by factors such as the physical and chemical properties of the encoded proteins, gene length, and CpG dinucleotide bias. Additionally, the researchers emphasized the significance of relative synonymous codon usage (RSCU) values in identifying patterns of synonymous codon usage. The researchers excluded certain codons, such as AUG and TGG, from their analyses to focus on the more relevant codons. Moreover, the findings suggested that the expression levels of MERS-CoV proteins varied between human and dromedary camel hosts, indicating a potential link between codon usage and the severity of diseases caused by the virus. The study concluded that optimizing MERS-CoV gene expression could enhance the understanding of the functional relevance of various proteins, potentially aiding in vaccine development. Overall, the research contributed valuable knowledge to the molecular evaluation of MERS-CoV, emphasizing the role of CUB in viral adaptation and evolution [118].

Numerous investigations on the genetic characteristics and adaptive processes of tomato brown rugose fruit virus (ToBRFV) were prompted by Ghorbani, A.'s earlier recognition of the virus as a serious danger to tomato output worldwide. As evidenced by the finding of more single nucleotide polymorphisms (SNPs) in the RNA-dependent RNA polymerase (RdRp) gene than in other isolates, the analysis carried out in this study verified that the Iranian isolate of ToBRFV exhibited unique genetic traits, raising the possibility of a new strain or sequencing artifact. The study also highlighted the importance of selection pressures and codon use bias (CUB), especially in the movement protein (MP) gene, which displayed the highest SNP density and transition/transversion bias, indicating possible evolutionary forces. These findings supported the adaptive modifications that increase the virus's ability to elude host defenses and were in line with earlier research showing the function of certain residues in the MP gene in avoiding host resistance mechanisms. Additionally, the phylogenetic analysis indicated that ToBRFV isolates were randomly distributed across the tree, suggesting extensive gene flow among populations globally, which was consistent with earlier studies that reported low genetic differentiation among geographic populations. The study's results provided evolution and adaptation, emphasizing the need for further research to develop effective control strategies against this economically impactful

virus. Overall, the literature underscored the complexity of ToBRFV's genetic landscape and the ongoing challenges it posed to agricultural sustainability [119].

Literature review of codon usage suggested that CUB is a significant phenomenon influenced by various evolutionary forces, including natural selection and mutation pressure. Studies have shown that different organisms exhibit unique codon preferences, which can be linked to their ecological adaptations and environmental conditions. For instance, research highlighted that species with similar phenotypic traits and habitats tend to share comparable codon preferences, indicating a relationship between codon usage and ecological factors. Furthermore, the effective number of codons (Nc) and other metrics have been employed to quantify CUB, revealing that variation in codon usage across different organisms is primarily driven by mutational forces, while variation within genomes is more influenced by selectional forces. This distinction is crucial for understanding the evolutionary dynamics of CUB. Additionally, studies have emphasized the role of GC content in shaping codon usage patterns, with findings suggesting that higher GC content correlates with specific codon preferences, further illustrating the complexity of codon usage in relation to evolutionary adaptations [120, 121].

CHAPTER 3

3. Research gap identification

Despite the significant potential and wide-ranging applications of *Staphylococcus* in fields such as healthcare, medicine, and the food industry, most prior studies have focused predominantly on a few specific pathogenic members of the genus. These studies have largely concentrated on the well-known species like *Staphylococcus aureus* and *Staphylococcus epidermidis*, which are responsible for a variety of infections in humans. However, there remains a notable gap in research concerning the broader *Staphylococcus* genus, particularly regarding the codon usage patterns across its diverse species. This lack of comprehensive genomic studies hinders a complete understanding of the evolutionary dynamics and genetic trends within the genus.

Investigating codon and amino acid usage across *Staphylococcus* species could provide novel insights into the genomic characteristics that distinguish different members of the genus. Such studies could reveal how these organisms adapt to various environmental and host-related factors, potentially opening new avenues for therapeutic interventions. For example, understanding the codon preferences and amino acid distribution may aid in the development of novel diagnostic tools or therapeutic candidates, specifically targeting *Staphylococcus* species. Additionally, such knowledge could be leveraged in optimizing industrial processes where *Staphylococcus* is used, such as in the production of fermented foods or biotechnology applications.

Therefore, given the aforementioned factors, the present study aims to comprehensively analyze and compare the codon and amino acid usage patterns across the entire *Staphylococcus* genus. By expanding our understanding of these genomic trends, this research seeks to contribute valuable insights that can inform both medical and industrial applications of *Staphylococcus* species.

AIM AND OBJECTIVE

3.1. Aim: Codon and Amino Acid Usage Analysis in Genus *Staphylococcus*

3.2. Objectives:

I. To inspect relative synonymous codon usage (RSCU) and relative amino acid usage (RAAU) patterns of the genomes of approximately 50 Staphylococcal species.

II. To investigate the extent of codon usage bias and contribution of factors like mutational bias and natural selection for translational efficiency.

III. To predict the codon adaptation index and deduce the optimal codons across all genomes followed by multivariate statistical analysis

IV. a) To compare codon context signatures across all genomes of *Staphylococcus* genus.

b) Experimental validation of codon usage bias in selected species

CHAPTER 4

4. Materials and methods

4.1. Data collection and genome annotation

The genomic data for this study was precisely curated from the esteemed Ensembl database (https://bacteria.ensembl.org/), a widely recognized repository of comprehensive genomic information. To ensure the integrity and reliability of the data, a rigorous refinement process was implemented. The accuracy of the analyses heavily relies on the quality and completeness of the genomic data used. Incomplete or low-quality genome assemblies could introduce biases and inaccuracies into the results. Thus, this process involved the elimination of sequences harboring ambiguous codons, internal stop codons, and those falling short of a minimum length threshold of 300 base pairs. This meticulous approach ensured that only high-quality, well-defined sequences were utilized for subsequent analyses. The adoption of procedural rigor was undertaken to anticipate and prevent potential inaccuracies in the sampling process. Sequences below this threshold may lack essential coding regions or contain fragmented data, which could compromise the accuracy and reliability of the analysis [88, 122]. The resultant compilation of coding sequences, is presented in a tabular format that delineates the quantity of sequences analyzed for each corresponding bacterial genome. There can be errors in gene annotation, particularly in identifying codons and determining their usage, could impact the reliability of the results. Misannotated genes or misidentification of start and stop codons could lead to inaccuracies in codon usage analyses.

4.2. Codon usage analysis tools and statistical measures

CodonW software version 1.4.2 was used to estimate the nucleotide compositional features in the bacterial genome, including the frequencies of Adenine (A), Guanine (G), Cytosine (C), and Thymine (T), as well as the makeup of G and C bases at the first (GC1), second (GC2), and third (GC3) codon positions, as well as the total GC and AT contents [123]. The software can be accessed at https://mybiosoftware.com/codonw-1-4-4-codon-usage-analysis.html. The RSCU (Relative Synonymous Codon Usage) for the relevant bacterial genomes was determined using the CodonW program. RSCU

serves as an index reflecting the diverse utilization of synonymous codons [124]. Assuming consistent synonymous codon use, it is measured as the ratio of a specific codon's observed frequency to its predicted frequency. The RSCU values for each genome were determined using the CodonW programme and were calculated as follows:

RSCU = Frequency of codon/ Expected frequency of codon (if codon usage was uniform)

ENc, also known as the effective number of codons, is a measure of codon usage bias that quantifies the overall redundancy in the usage of synonymous codons (codons that code for the same amino acid) [125]. The CodonW program was used to determine the effective number of codons (ENc) scores for the bacterial genomes in order to measure the level of bias in codon use. An amino acid expressed by a single codon is said to have significant codon use bias (ENc = 20), whereas an amino acid translated by all of its synonymous codons is said to have no codon bias (ENc = 61). While lower values indicate a stronger preference for certain codons, higher ENc scores indicate a more equal distribution of synonymous codon use [126, 127] . The following formula is used to determine ENc:

$$ENc = 2 + 9/F_2 + 1/F_3 + 5/F_4 + 3/F_6$$

F denotes the probability that two randomly selected codons would encode the same amino acid, whereas Fk reflects the average value of k-fold degenerate amino acids (where k = 2, 3, or 4) [128]. The bacterial genomes' ENc scores offer important information on the general patterns of codon use and their possible effects on gene expression and evolutionary dynamics [129].

The Grand Average Hydropathicity score (**GRAVY**) and Relative Amino Acid Utilization (**RAAU**) were computed using the CodonW program in order to describe the amino acid contents of the poorly studied bacterial genomes. [130]**,** and aromaticity **(Aromo**) [131]. RAAU quantifies the relative abundance of each amino acid in

bacterial genomes, offering insights into the overall amino acid composition of the proteins encoded by these genomes. GRAVY, which assesses a protein's overall hydrophobicity, was determined by evaluating the hydropathicity scores of the individual amino acids. This measure provides information about the potential solubility and membrane localization of the proteins encoded by these genomes. The calibration used to calculate the GRAVY score is as follows:

$$\text{GRAVY} = \frac{1}{N}\sum_{i=1}^{N} k_i$$

N is the total number of amino acids, and k_i is the hydrophobicity index of the i^{th}amino acid .

Aromo, which quantifies a protein's aromaticity, was calculated to evaluate the relative abundance of aromatic amino acids (phenylalanine, tryptophan, and tyrosine) in the proteins. Aromatic residues are vital for protein structure, function, and their interactions with other molecules [132]. Aromo is written as:

$$\text{Aromo} = \frac{1}{N}\sum_{i=1}^{N} v_i$$

In this equation, N represents the total number of amino acids in the protein sequence, and v_i is set to 1 for aromatic amino acids and 0 for non-aromatic amino acids.

By employing CodonW software, a comprehensive assessment of the amino acid composition of the under-researched bacterial genomes was achieved, providing valuable insights into their proteomes and potential biological functions [91].

Some of the analyses, such as ENc and CAI, rely on specific statistical assumptions. Violations of these assumptions could lead to biased results. It's essential to validate the assumptions underlying each analysis method.

4.3. Neutrality plot

To assess the relative contributions of mutation pressure and natural selection in shaping codon usage patterns, neutrality plot analysis was employed. This method involves plotting the GC3 values (GC content at the third codon position) of bacterial genes on the x-axis against their corresponding GC12 values (average GC content at the first two codon positions) on the y-axis [133]. The slope of the resulting regression line provides insights into the level of compositional constraint acting on the genes. A steeper slope indicates stronger compositional constraint, suggesting that natural selection plays a more significant role in shaping codon usage patterns. Conversely, a shallower slope indicates weaker compositional constraint, implying that mutation pressure has a greater influence on codon usage [134].

Caution is crucial in interpreting neutrality plots, as misinterpretation can lead to erroneous results. These plots offer insights into the balance between mutation and selection forces shaping codon usage bias. A clear understanding of their implications ensures accurate conclusions about the evolutionary dynamics of genomic sequences.

4.4. Assessment of translational selection

Translational selection, also known as translational efficiency, is a process by which cells favor the use of codons that are recognized more efficiently by their cognate transfer RNAs (tRNAs). This occurs because the efficiency of translation can be affected by the availability of tRNAs and the speed at which they can bind to codons on the mRNA. Translational selection is measured by the P2 index, which is a numerical value that ranges from 0 to 1 [117, 135]. A P2 value of 1 indicates that all codons are used equally efficiently, while a P2 value of 0 indicates that the most frequently used codons are also the most efficiently translated. Translational selection is an important factor in gene expression because it can affect the rate at which a gene is translated into protein. Genes with higher P2 values are generally translated more efficiently than genes with lower P2 values. This can have a significant impact on the abundance and function of the proteins encoded by these genes. Translational selection (P2) measures the degree of accuracy in codon-anticodon interaction and is a significant parameter for

evaluating the impact of translational selection on a gene or set of genes. P2 is calculated as:

$$P2 = \frac{WWC + SSU}{WWY + SSY}$$

In this case, W stands for either adenine (A) or thymine (T), S for either cytosine (C) or guanine (G), and Y for either cytosine (C) or thymine (T). Natural selection significantly affects translation when the P2 value is higher than 0.50 [136].

4.5. Codon adaptation index

The codon adaptation index (CAI) is used as a key metric to predict gene expression levels by comparing codon use patterns between relevant genes and a group of highly expressed genes. Higher values indicate a closer relation in expression patterns to the reference set. The generated values range from 0 to 1. The reference collection for this work consisted of highly expressed ribosomal protein coding genes, and the DAMBE tool was used to calculate CAI values for bacterial genomes [137, 138].

4.6. Correspondence analysis

To comprehensively analyse the variations in codon and amino acid usage patterns within the investigated bacterial genomes, Correspondence Analysis (CoA), a multivariate statistical technique, was employed [76, 139]. This method effectively visualizes the underlying relationships among variables by projecting them onto a set of orthogonal axes, with each axis capturing a diminishing proportion of the observed variance. The RSCU (Relative Synonymous Codon Usage) and RAAU (Relative Amino Acid Usage) data extracted from the bacterial genomes served as input for the CoA analysis, which was performed using the CodonW software [140, 141].

4.7. Analysis of relative abundance of dinucleotides

An internal BioPerl script was used to compute the observed and predicted frequencies of dinucleotides in order to ascertain the relative abundance of these segments throughout the bacterial genomes. [142, 143]. The odds ratio was then calculated as follows:

$$P_{xy} = \frac{f_{xy}}{f_x f_y}$$

The dinucleotide XY's observed and predicted frequencies are shown by fxy and fxfy, respectively, whilst the observed frequencies of nucleotides X and Y are represented by fx and fy, respectively. Over-represented dinucleotides are those with Pxy values > 1.25, while those with Pxy values < 0.78 are considered under-represented [144, 145].

4.8. Estimation of relative synonymous codon pair usage and codon pair score

The Relative Synonymous Codon Pair Usage (RSCPU), frequently employed to assess the ratio of observed to expected frequencies of specific codon pairs, was utilized in this study [146, 147]. RSCPU values for the codon pairs were calculated through the application of a custom BioPerl script developed in-house. wRSCU was calculated using following equation:

$$\text{RSCPU} = \frac{\text{Observed frequency of a codon pair}}{\text{Expected frequency of the codon pair}}$$

The natural logarithm of RSCPU was calculated to obtain the codon pair score (CPS). Over-represented codon pairs were those having positive CPS ratings, whereas under-represented codon pairs had negative CPS scores [148].

While codon usage patterns can provide insights into gene expression and adaptation, attributing specific biological functions solely based on codon bias can be challenging. Functional validation through experimental studies is necessary to confirm hypotheses generated from computational analyses.

4.9. Statistical analysis

The indices for Codon usage were systematically generated and thoroughly verified using a suite of statistical methods to ensure the findings were both accurate and reliable. IBM SPSS Statistics (Version 28) served as the primary tool for conducting all statistical analyses, leveraging its powerful features for data management, exploration, and hypothesis testing. The analysis process likely began with data cleaning and

preparation, which involved identifying and handling any missing values or outliers to prevent biases in the results. Descriptive statistics were then used to summarize the central tendencies and variability in the data, providing an initial overview of the distribution and trends in codon usage. Subsequently, inferential statistical tests, including t-tests, ANOVA, or chi-square tests, were likely performed to analyze relationships and differences within the dataset. For instance, t-tests or ANOVA could have been used to compare codon usage across different groups or conditions, while correlation or regression analyses could have assessed the strength and direction of associations between codon usage indices and other relevant variables [149-151].

4.10. Experimental validation of codon usage bias in selected species by genome sequencing

4.10.1. Preparation of Cultures

Freeze-dried cultures of *Staphylococcus hominis* (strain 4435) and *Staphylococcus epidermidis* (strain 6152) were procured and revived following standard protocols. Nutrient Agar, which was made by combining 1.0 g beef extract, 2.0 g yeast extract, 5.0 g peptone, 5.0 g NaCl, and 15.0 g agar in 1.0 L of distilled water, was used to cultivate *S. hominis*. After being autoclaved and allowed to cool to 45°C, the mixture was transferred into sterile petri plates. Tryptic Soy Agar (TSA), which was made by dissolving 30.0 g of Trypticase Soy Broth and 15.0 g of agar in 1.0 L of distilled water, was used to cultivate *S. epidermidis*. The mixture was autoclaved, cooled, and poured as with Nutrient Agar, and both species were incubated at 37°C for 24 hours to ensure optimal growth [152-154].

4.10.2. DNA Extraction

Following incubation, bacterial colonies from each medium were harvested for genomic DNA extraction. DNA isolation was carried out using the DNeasy Ultraclean Microbial Kit (Qiagen, Cat. No. 12224-250). DNA was eluted in 50 µL buffer, and its quality and concentration were assessed using a Nanodrop1000 spectrophotometer. Two readings were taken per sample to confirm concentration and purity, with an

A260/280 ratio close to 1.8 considered optimal. DNA samples that met quality criteria were then prepared for sequencing [155-157].

4.10.3. Genome Sequencing

Genome sequencing quality was evaluated by processing raw reads with FastQC v.0.11.9, followed by preprocessing using Fastp v.0.23.4. Fastp parameters included a minimum length requirement of 50, poly-G trimming, and quality filtering at Phred 30, ensuring high-quality reads for analysis. Taxonomic profiling of the pre-processed reads was conducted using Kraken2 v.2.1.2 with the KRAKEN2 bacterial database, and results were visualized using Krona for clarity. To confirm the species, the processed reads were aligned with the reference genomes from NCBI RefSeq for *S. hominis* (GCF_003812505.1) and *S. intermedius* (GCF_002374235.1) using HoCoRT. After mapping with HoCoRT, contaminant-free reads that matched the reference genome were selected for assembly [158].

De novo genome assembly was performed with Unicycler v.0.4.4 under default parameters to generate a draft genome, and completeness was verified using CheckM2 v.1.0.1. QUAST v.5.0.2 was employed to assess the quality of the genome assembly, while extracted 16S rRNA sequences from the assembly were analyzed using ContEST16S (EzBioCloud) for further taxonomic verification. Comprehensive genome annotation was carried out with BAKTA, referencing various functional and resistance gene databases (e.g., AMRFinderPlus, COG, ISFinder, Pfam). Antimicrobial resistance gene prediction was completed with CARD-RGI v.6.0.3, and virulence factor detection was performed using PathoFact v.1.0 [158-160].

CHAPTER 5

5. Results and Discussion:

5.1. AT-rich *Staphylococcal* genomes elegantly shape the selection of preferred codons

Brocchieri *et al*. proposed that the G-C content within bacterial genomes varies significantly, with reported values ranging from 17% to 75% [161]. This wide range highlights the substantial genetic diversity observed among bacterial species. Additionally, Rocha *et al*. have pointed out that bacteria inhabiting free-living environments tend to possess higher G-C content compared to those that reside within host organisms. The elevated G-C content in free-living bacteria is attributed to the availability of G-C-rich metabolites in nutrient-rich environments, which support efficient genomic composition. In contrast, bacteria living within host organisms are more exposed to AT-rich metabolites, such as ATP, which are more prevalent in host environments. This metabolic difference drives the selection of AT-rich genomes in pathogenic bacteria, as the lower synthesis cost of AT-rich genomes allows these organisms to replicate more rapidly and adapt more effectively to the host environment [162].

Understanding the biological significance of genomic G-C content in Staphylococcus provides critical insights into the evolutionary dynamics, pathogenicity, and adaptation strategies of these bacteria. The genomic composition plays a vital role in shaping the overall fitness of bacteria and influencing interactions between host and pathogen. Through our investigation, we observed significant variations in the genomic G-C content across 48 species within the Staphylococcus genus, ranging from 34.27% to 36%. This variation reflects the adaptive mechanisms that these bacteria employ in response to environmental pressures and host conditions. Specifically, the G-C content of coagulase-positive species was found to be 36%, with a standard error mean of ±0.00879%, indicating a relatively consistent G-C content in these species. In comparison, coagulase-negative species exhibited G-C content values of 34.47% (±0.00305%) and 34.29% (±0.00222%) for coagulase variables, highlighting a slightly lower G-C content in these species.

The observed variation in G-C content also plays a crucial role in influencing the selection of preferred codons across the 48 members of the Staphylococcus genus. Our study revealed that the choice of preferred codons was closely linked to the genomic G-C content. Coagulase-positive species (CoPS) demonstrated comparatively higher utilization of G-C rich preferred codons, with 28% of their preferred codons being G-C rich. In contrast, coagulase-negative species (CoNS) and coagulase variables displayed lower usage rates of G-C rich preferred codons, with 26.9% and 26.08%, respectively.

Furthermore, our analysis showed a marked preference for A/T at the third nucleotide position in the preferred codons of both coagulase-positive and coagulase-negative species. Coagulase-positive species exhibited a 72% preference for A/T at this position, while coagulase-negative species and coagulase variables showed preferences of 73.07% and 73.91%, respectively **(Table 1a-c).** The optimal codons were predominantly A/T-rich in composition, particularly favoring A/T at the wobble position. This bias toward an AT-rich genomic structure highlights the pathogen's ability to survive within the host environment, where AT-based metabolites are readily available. A genome enriched in AT content enables the pathogen to efficiently utilize these metabolites for growth and sustenance post-infection.

Moreover, the less stable nature of AT-rich regions within the genome plays an additional role in the pathogen's replication process. These AT-rich regions function as crucial sites for DNA unwinding, facilitating the initiation of replication. By having a genome that is predominantly AT-rich, pathogenic bacteria are better equipped to initiate and sustain rapid replication within host organisms, contributing to their overall fitness and survival. This adaptive strategy underscores the importance of genomic composition in shaping the evolutionary success of pathogenic bacteria of Staphylococcus genus.

Table 1a: **Characteristics of the coagulase positive species of *Staphylococcus* genomes**

Organism	Accession number	GC%	Number of protein Coding Genes	Gene Transcript	P2	T3	C3	A3	G3
Staphylococcus aureus	GCA_001049575	32.59	2,861	2,935	0.385282	0.498369	0.136584	0.522306	0.140395
Staphylococcus simiae CCM 7213 = CCUG 51256	GCA_000235645	32.76	2,592	2,648	0.382787	0.517616	0.12729	0.532933	0.110871
Staphylococcus argenteus str. 58113	GCA_003967115	35.98	2,677	2,756	0.383444	0.502301	0.129222	0.53317	0.130003
Staphylococcus pseudintermedius	GCA_004794825	37.63	2,401	2,528	0.434924	0.449213	0.185372	0.455878	0.194592
Staphylococcus intermedius NCTC 11048	GCA_900458545	38.03	2,805	2,886	0.430196	0.449877	0.184571	0.467299	0.190661
Staphylococcus delphini	GCA_002369735	37.56	2,372	2,508	0.443639	0.4367	0.197152	0.448717	0.202142
Staphylococcus lutrae	GCA_002902165	34.39	2,238	2,354	0.436353	0.448971	0.178191	0.453385	0.203092

Table 1b: Characteristics of the co-variables of *Staphylococcus* genomes

Organism	Accession number	GC%	Number of protein coding Genes	Gene Transcript	P2	T3	C3	A3	G3
Staphylococcus hyicus	GCA_003580585	36.5	2,306	2,413	0.41954	0.47403	0.162411	0.485781	0.163569
Staphylococcus agnetis	GCA_003040835	32.05	2,300	2,427	0.4213	0.475782	0.164207	0.481936	0.160185

Table 1c: Characteristics of the coagulase negative species of *Staphylococcus* genomes

Organism	Accession number	GC%	Number of proteinCoding Genes	Gene Transcript	P2	T3	C3	A3	G3
Staphylococcus epidermidis	GCA_004329135	32	2,728	2,918	0.382128	0.51875	0.131865	0.52159	0.123643
Staphylococcus haemolyticus	GCA_003580795	32.7	3,722	4,014	0.403959	0.495201	0.152707	0.508703	0.13255
Staphylococcus hominis	GCA_004329095	31.4	3,231	3,433	0.374345	0.529776	0.122008	0.533314	0.110409
Staphylococcus capitis	GCA_004329465	32.8	3,257	3,457	0.389569	0.515859	0.142359	0.512223	0.121188
Staphylococcus pettenkoferi	GCA_002884615	38.8	2,345	2,469	0.450035	0.439543	0.230062	0.421142	0.201602
Staphylococcus simulans	GCA_003043455	36	2,370	2,477	0.428576	0.450482	0.194546	0.492703	0.145808
Staphylococcus warneri	GCA_003043215	32.6	2,368	2,478	0.393085	0.499223	0.144441	0.533305	0.116295
Staphylococcus auricularis	GCA_002902455	37.2	2,004	2,146	0.442582	0.437778	0.216853	0.475886	0.156217
Staphylococcus devriesei	GCA_002902625	33.3	2,223	2,392	0.398609	0.501603	0.156089	0.510088	0.122831
Staphylococcus pasteuri	GCA_002276895	31.5	2,714	2,847	0.393983	0.487076	0.154592	0.517645	0.12675
Staphylococcus	GCA_003040195	36.6	2,377	2,518	0.429337	0.463475	0.184398	0.472798	0.166523

chromogenes									
Staphylococcus saccharolyticus	GCA_003970495	32	2,308	2,431	0.378139	0.518576	0.129	0.548707	0.095184
Staphylococcus edaphicus str. CCM 8730	GCA_002614725	33.35	2,554	2,690	0.395539	0.488024	0.143266	0.520248	0.137359
Staphylococcus petrasii subsp. croceilyticus	GCA_002902575	33.3	2,229	2,346	0.402802	0.496422	0.160391	0.513035	0.111497
Staphylococcus massiliensis S46	GCA_000314555	36.5	2,324	2,386	0.425174	0.473151	0.196347	0.454775	0.169472
Staphylococcus schleiferi	GCA_900458895	35.9	5,124	5,283	0.405405	0.489528	0.150442	0.500714	0.149311
Staphylococcus condimenti	GCA_004209905	34.63	2,934	3,115	0.412302	0.474121	0.17382	0.495943	0.141743
Staphylococcus piscifermentans	GCA_003970515	36.6	2,536	2,654	0.394507	0.491295	0.143144	0.525364	0.127466
Staphylococcus carnosus subsp. carnosus	GCA_003970565	34.535	2,323	2,462	0.396626	0.505628	0.153812	0.51097	0.118514
Staphylococcus arlettae	GCA_900457375	33.4	2,594	2,677	0.39834	0.492536	0.141009	0.523243	0.128061

Staphylococcus kloosii	GCA_001593625	32.8	2,484	2,568	0.394373	0.493541	0.149068	0.525544	0.118084
Staphylococcus muscae	GCA_900187005	37.35	1,985	2,064	0.435799	0.445336	0.176238	0.479933	0.179325
Staphylococcus rostri	GCA_002902145	38.4	2,219	2,342	0.444803	0.429407	0.19625	0.456286	0.199577
Staphylococcus microti	GCA_900458705	38	2,334	2,412	0.439106	0.438734	0.185915	0.46252	0.195775
Staphylococcus felis	GCA_003388115	34.9	2,780	2,924	0.410212	0.483925	0.153713	0.493631	0.156718
Staphylococcus caprae M23864:W1	GCA_004329095	33.5	2,563	2,620	0.397848	0.498437	0.157791	0.508561	0.132013
Staphylococcus saprophyticus	GCA_900458845	33	2,579	2,660	0.385699	0.499232	0.134384	0.531569	0.123978
Staphylococcus gallinarum	GCA_003577565	33.1	3,176	3,310	0.395537	0.487226	0.138757	0.5197	0.142286
Staphylococcus equorum	GCA_001747785	33	2,915	3,022	0.39345	0.486697	0.146952	0.529796	0.13184
Staphylococcus xylosus	GCA_003578865	32.7	2,824	2,953	0.386062	0.49845	0.135887	0.529381	0.128064
Staphylococcus succinus	GCA_003041455	32.9	3,604	3,787	0.410503	0.463804	0.172313	0.470266	0.172296
Staphylococcus nepalensis	GCA_003042875	33.018	3,823	4,050	0.405771	0.478383	0.159238	0.513627	0.1335
Staphylococcus pseudoxylosus	GCA_003697915	32.977	2,898	3,013	0.393897	0.491438	0.142195	0.524807	0.130966

str. S04009									
Staphylococcus cohnii	GCA_005861955	32.4	2,421	2,592	0.389407	0.49816	0.136586	0.52898	0.12486
Staphylococcus lugdunensis	GCA_004329485	33.7	2,835	3,000	0.398676	0.499611	0.143369	0.51001	0.135722
Staphylococcus sciuri	GCA_003041975	32.5	3,442	3,611	0.386728	0.490873	0.136896	0.545667	0.117546
Staphylococcus lentus	GCA_001651255	31.8	2,484	2,609	0.37871	0.493016	0.129868	0.557838	0.110079
Staphylococcus fleurettii	GCA_003577995	31.6	3,089	3,256	0.377953	0.497226	0.125075	0.560047	0.109619
Staphylococcus vitulinus	GCA_003043315	32.547	2,520	2,643	0.382998	0.489296	0.136565	0.544829	0.12116

5.2. ENc influencing codon usage bias

According to Nasrullah *et al*., this study plotted GC3s, or the GC content at the third codon position of synonymous codons, against the effective number of codons (ENc) in order to investigate the combined effects of mutational pressure and natural selection on codon usage bias throughout the genome [163]. The ENc metric is an effective measure for quantifying codon usage bias by determining the extent to which synonymous codons deviate from a random, equal usage. A lower ENc value signifies stronger codon bias, indicating that specific codons are preferred in the genome. This deviation is of evolutionary interest, as it can reveal the forces acting on a genome—particularly, how natural selection optimizes certain codons for increased translational efficiency, accuracy, or protein folding, conferring selective advantages, especially in highly expressed genes or specific environmental contexts.

ENc provides key insights into the balance between two opposing forces: mutational pressure, which introduces random alterations in DNA sequences, and natural selection,

which favors codons that enhance translation efficiency or accuracy. A lower ENc suggests that natural selection is favoring certain codons to optimize protein synthesis, potentially by improving ribosomal speed or reducing translation errors. For instance, a genome with an ENc of 20 shows extreme codon bias, with only 20 codons being used for the 20 amino acids, reflecting strong selective pressure on codon choice. This may be driven by the need for codons that align with abundant tRNAs, boosting translational efficiency. In contrast, an ENc of 61 implies no codon usage bias, where all 61 codons are equally utilized to encode the 20 amino acids, likely due to minimal selective pressure and a dominance of mutational pressure leading to random codon distribution. Genomes with ENc values below 35 exhibit strong codon usage bias, where natural selection plays a significant role in shaping codon patterns, likely driven by factors such as tRNA availability and the need for rapid protein synthesis. On the other hand, genomes with ENc values above 35 show more relaxed codon usage, indicating weaker selective forces and a more equal usage of synonymous codons. The ENc values observed in this study present a broad overview of how different genomes manage the opposing influences of selection and mutation, with the ENc-GC3 plot visualizing these trends. Higher ENc values suggest a more random codon usage influenced by mutational pressure, while lower values indicate stronger selective forces driving a preference for specific codons. In the analysis of the GC3s versus ENc plot, the positions of genes highlight the predominant forces shaping codon usage. Our findings revealed significant codon usage bias among CoPS (coagulase-positive staphylococci), CoNS (coagulase-negative staphylococci), and co-variables, with bias decreasing in the order of co-variables > CoPS > CoNS. The ENc values for the three Staphylococcus groups were slightly elevated, with co-variables showing an average ENc of 43.41, CoPS 42.43, and CoNS 41.51. These elevated ENc values indicate a more liberal codon usage pattern with reduced bias among synonymous codons. The GC3s vs ENc plot emphasizes the role of translational selection in shaping codon usage patterns within the Staphylococcus genus (**Fig. 2A-2F**). This preference for specific codons implies a lower tRNA burden, which may streamline the pathogen's genomic architecture and result in fewer genes. Consistent with this observation, CoPS has an average of 2564 genes, while CoNS and co-variables have slightly fewer, with 2750 and 2303 genes, respectively.

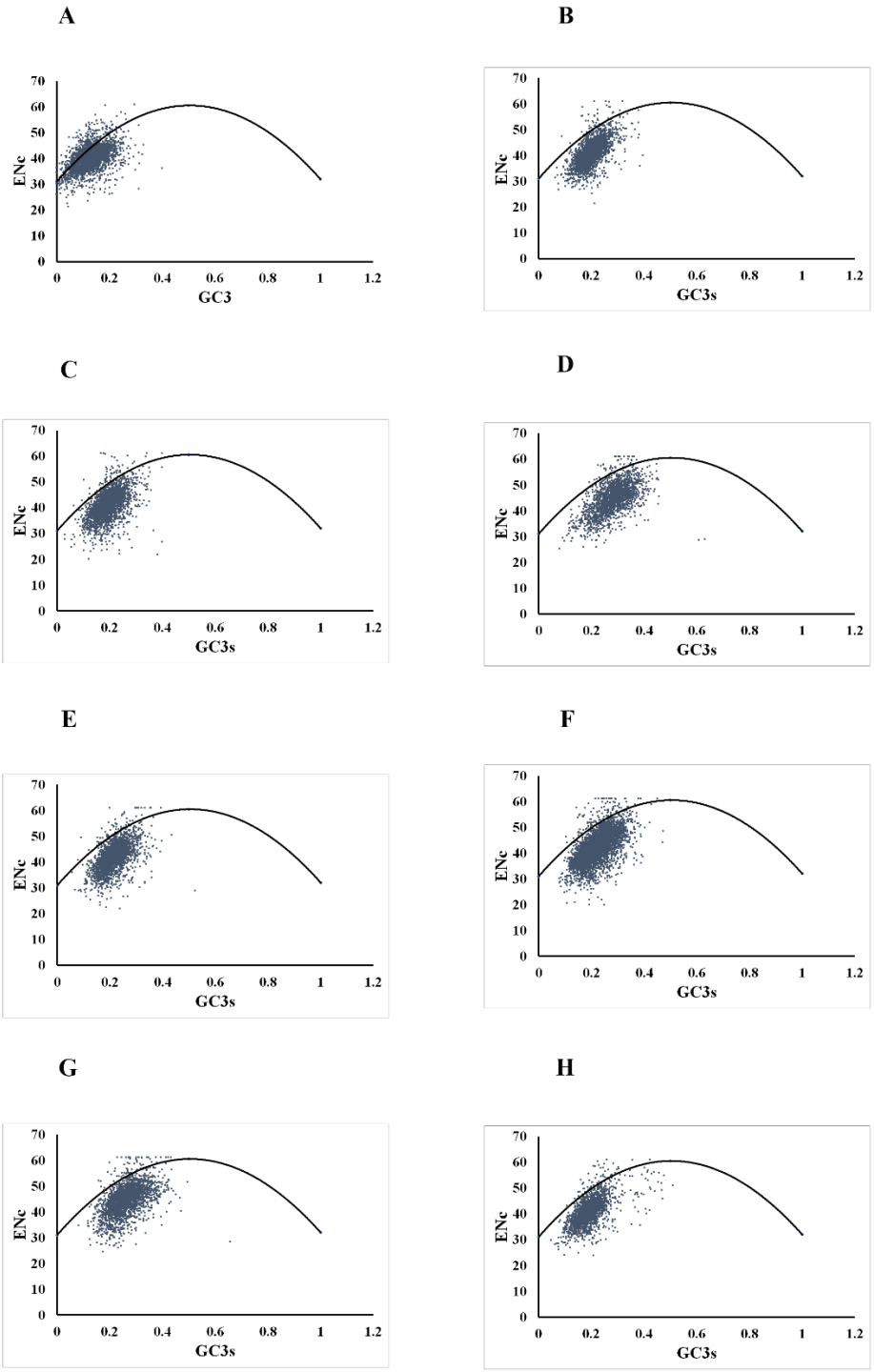


Figure 2A: GC3 *vs* Enc plot, with each data point representing an individual gene A). *Staphylococcus argenteus str. 58113* B). *Staphylococcus arlettae* C). *Staphylococcus auricularis* D). *Staphylococcus capitis* E). *Staphylococcus caprae M23864:W1* F). *Staphylococcus carnosus subsp. carnosus* G). *Staphylococcus chromogenes* H). *Staphylococcus cohnii*

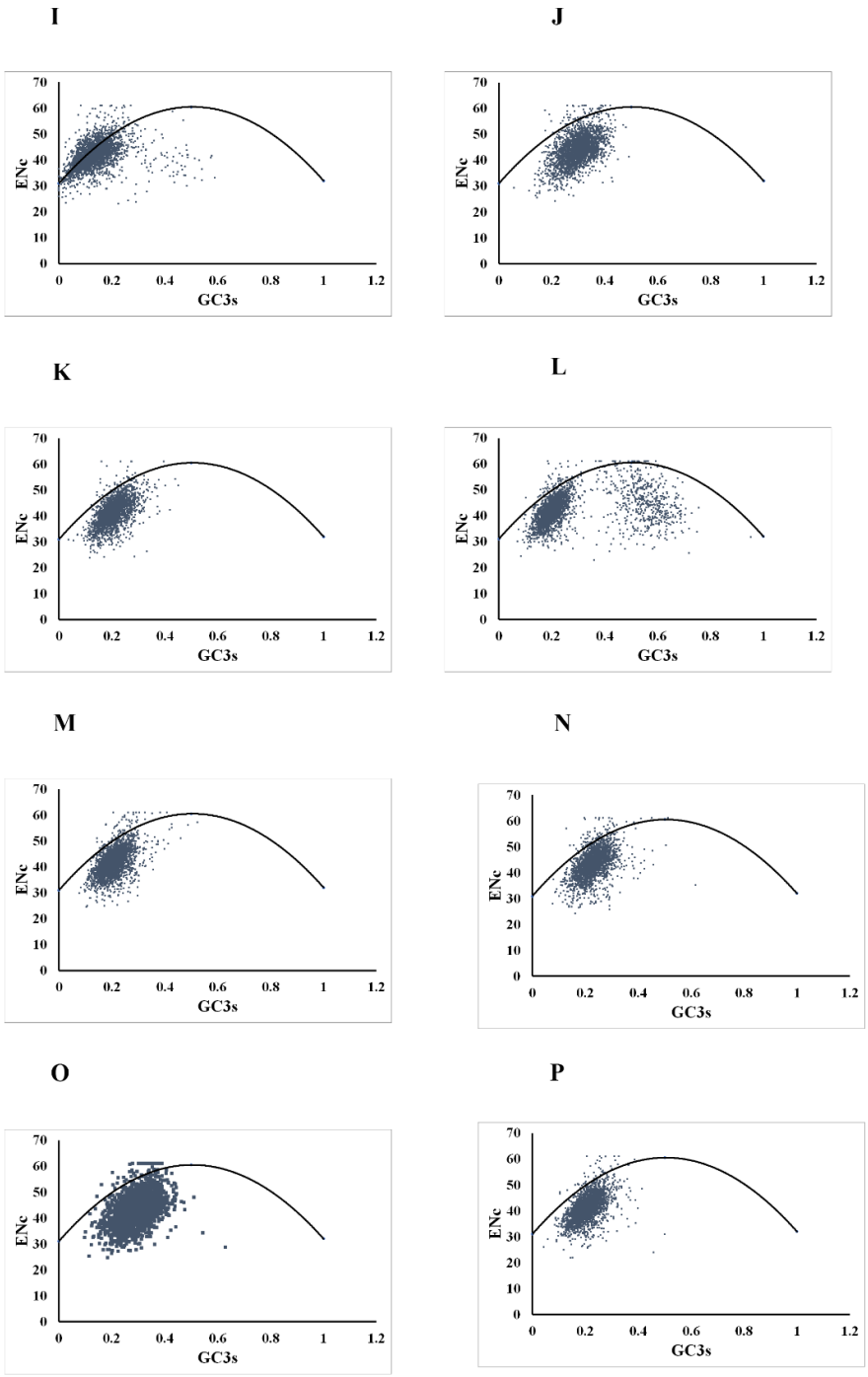


Figure 2B: GC3 *vs* Enc plot, with each data point representing an individual gene I). *Staphylococcus condimentistr. 58113* J). *Staphylococcus devriesei* K). *Staphylococcus edaphicus str. CCM 8730* L). *Staphylococcus equorum* M). *Staphylococcus felis M23864:W1* N). *Staphylococcus felis* O). *Staphylococcus fleurettii* P). *Staphylococcus gallinarum*

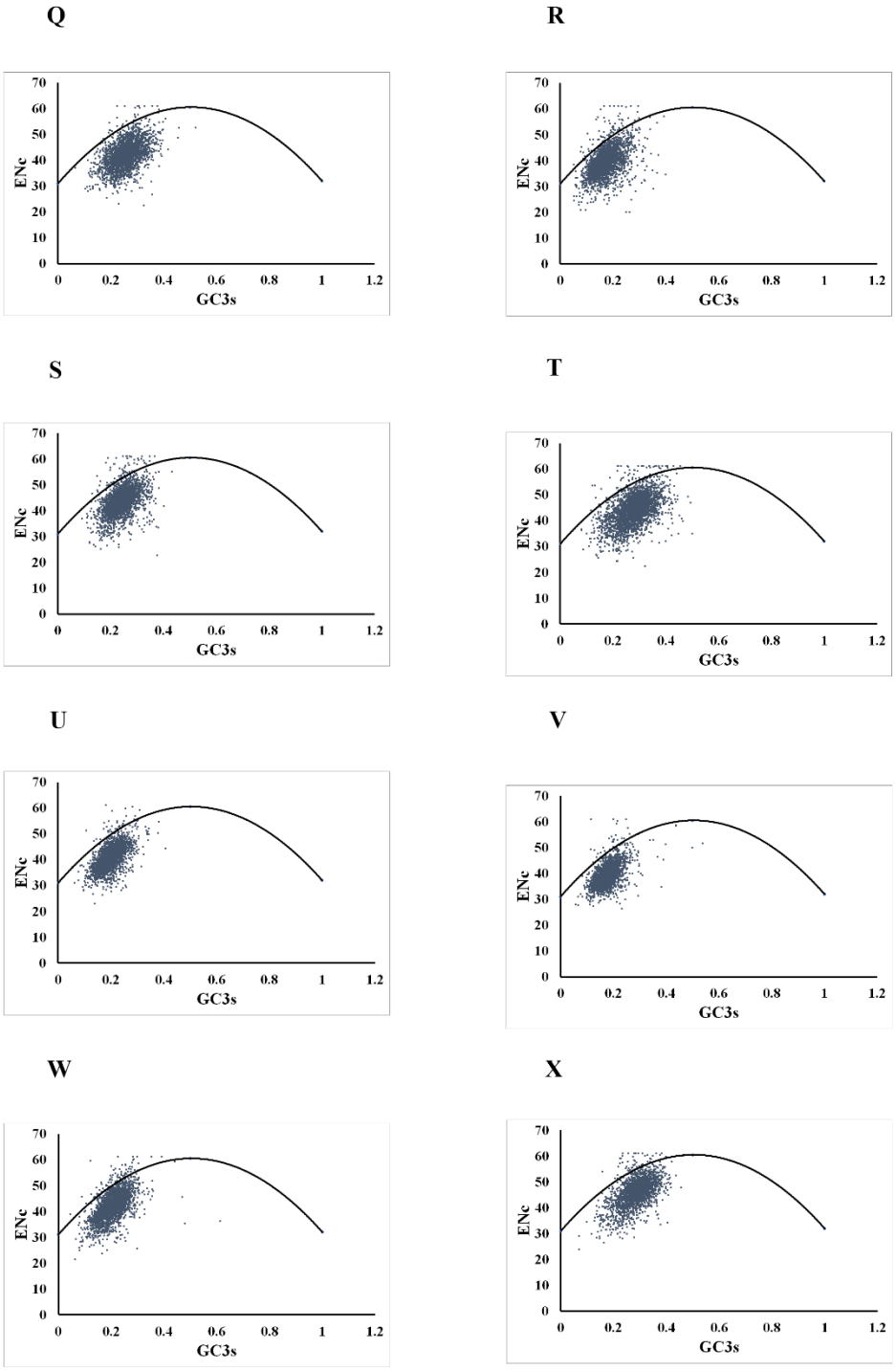


Figure 2C: GC3 *vs* Enc plot, with each data point representing an individual gene Q). *Staphylococcus haemolyticus* R). *Staphylococcus hominis* S). *Staphylococcus hyicus* T). *Staphylococcus intermedius NCTC 11048* U). *Staphylococcus kloosii* V). *Staphylococcus lentus* W). *Staphylococcus lugdunensis* X). *Staphylococcus lutrae*

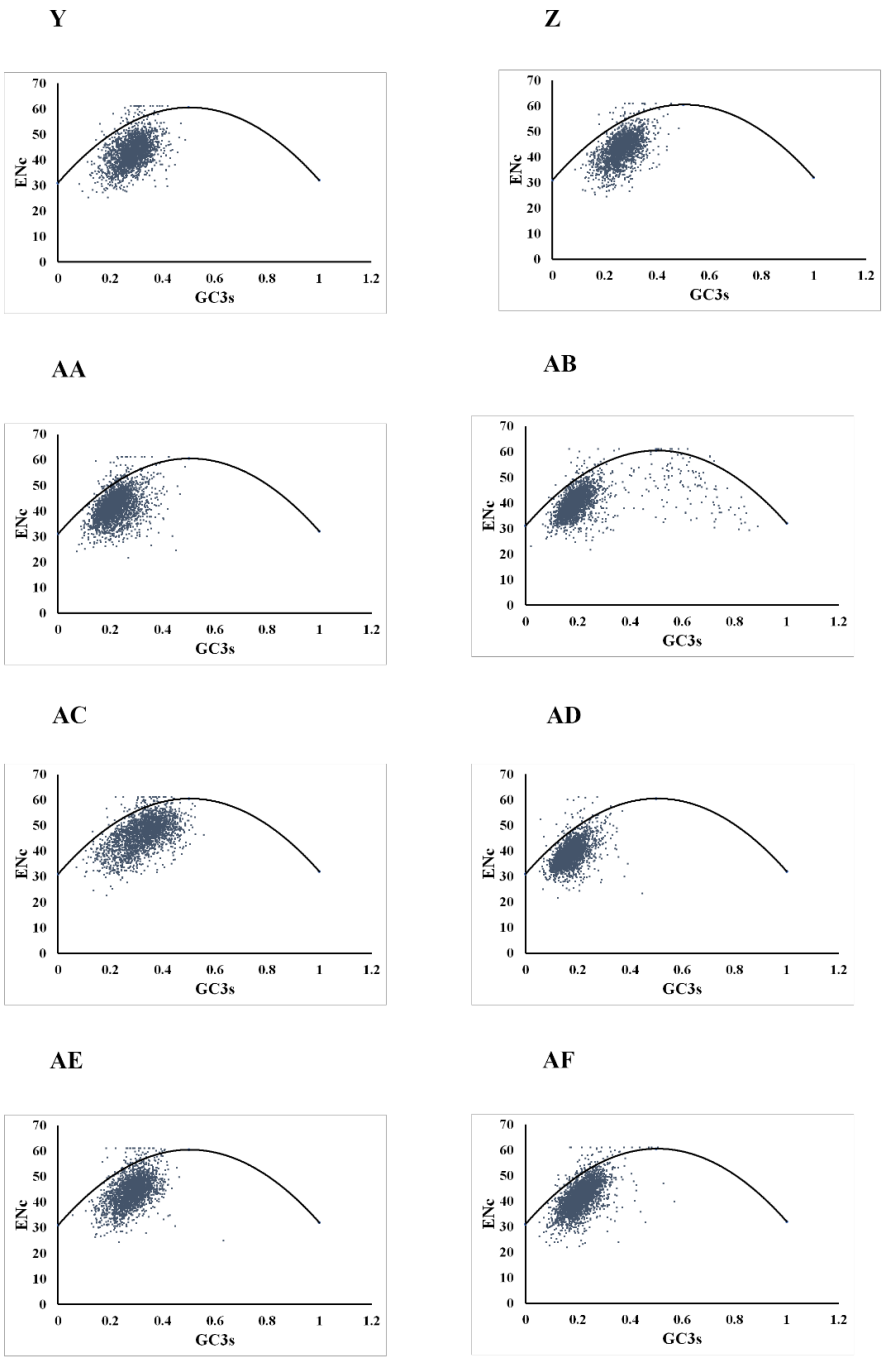


Figure 2D: GC3 *vs* Enc plot, with each data point representing an individual gene Y). *Staphylococcus massiliensis S46* Z). *Staphylococcus microti* AA). *Staphylococcus muscae* AB). *Staphylococcus nepalensis* AC). *Staphylococcus pasteuri* AD). *Staphylococcus petrasii subsp. croceilyticus* AE). *Staphylococcus pettenkoferi* AF). *Staphylococcus piscifermentans*

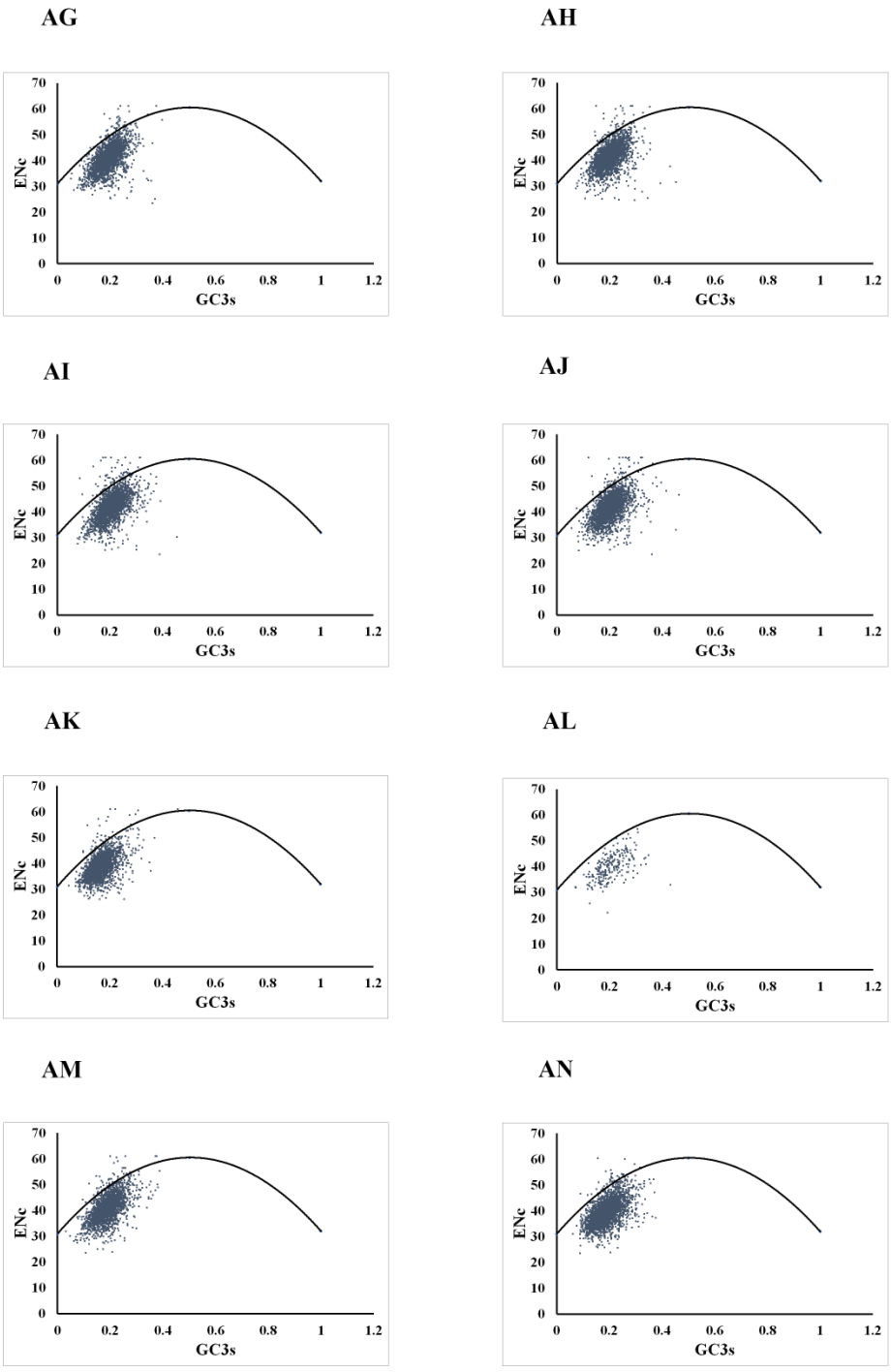


Figure 2E: GC3 *vs* Enc plot, with each data point representing an individual gene AG). *Staphylococcus pseudintermedius* AH). *Staphylococcus pseudoxylosus str. S04009* AI). *Staphylococcus rostri* AJ). *Staphylococcus saccharolyticus* AK). *Staphylococcus saprophyticus* AL). *Staphylococcus schleiferi* AM). *Staphylococcus sciuri* AN). *Staphylococcus simiae CCM 7213 = CCUG 51256*

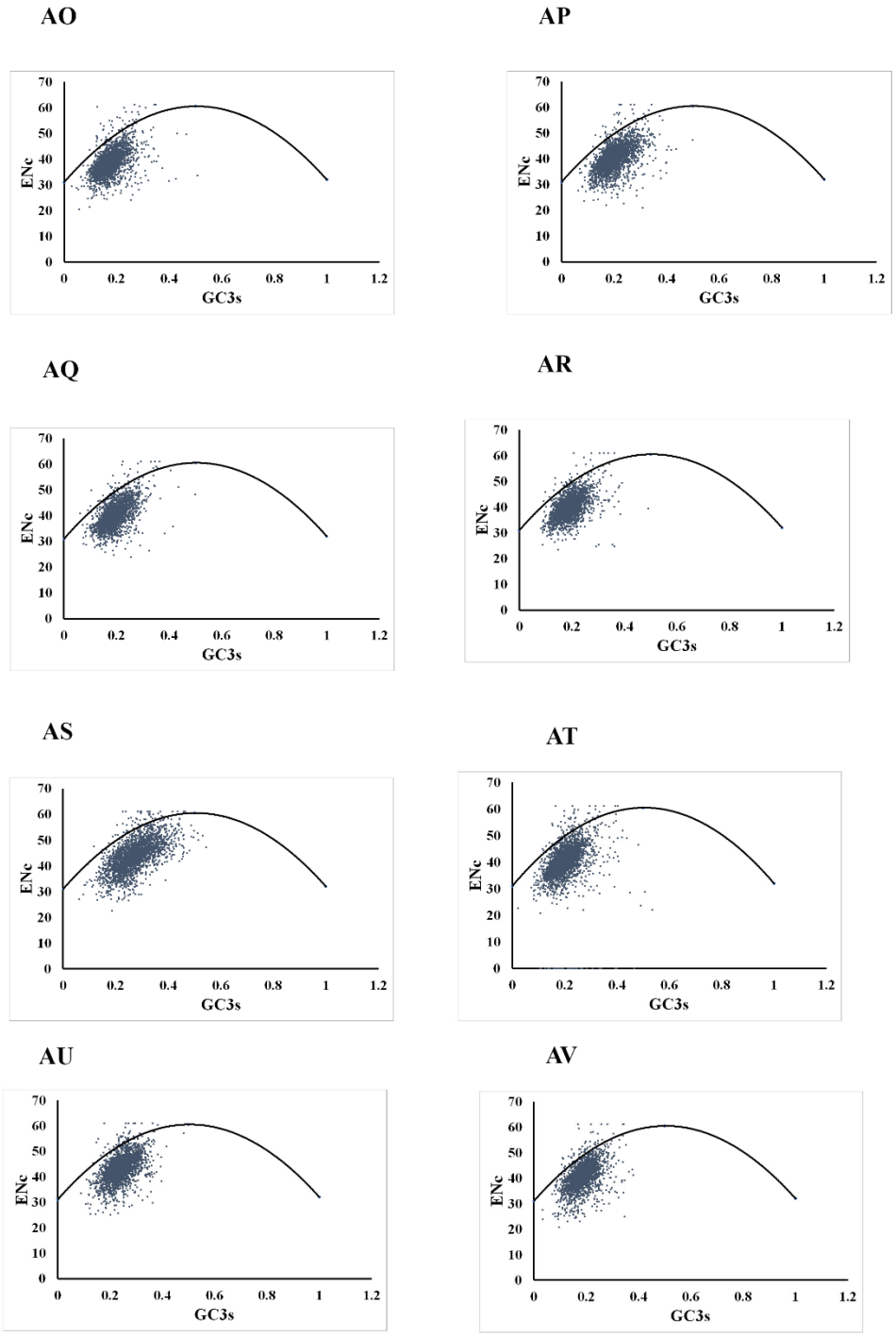


Figure 2F: GC3 *vs* Enc plot, with each data point representing an individual gene AO). *Staphylococcus simulans* AP). *Staphylococcus succinus* AQ). *Staphylococcus vitulinus* AR). *Staphylococcus warneri* AS). Staphylococcus xylosus AT) *Staphylococcus aureus* AU) *Staphylococcus agnetis* AV) *Staphylococcus epidermidis*

5.3. Neutrality Plot: Deciphering Selection Forces in *Staphylococcus* Genome

Following Duan and Antezana's methodology, a neutrality plot was created between the GC content at the third codon position of synonymous codons (GC3s) and the GC content at the first and second codon positions (GC12) in order to evaluate the impact of mutational pressure on codon usage throughout the genome [164]. The neutrality plot is a widely used graphical tool for evaluating the relative contributions of mutational pressure and natural selection in shaping the nucleotide composition of protein-coding genes. By plotting GC3s against GC12, this approach allows researchers to examine whether mutational bias or selective pressures have a more significant role in determining codon usage patterns. The fundamental premise of this plot is that if mutational pressure is the primary driver of codon usage, there should be a strong correlation between GC3s and GC12, reflected by a regression coefficient close to 1. Conversely, if selective forces dominate, particularly those related to translational efficiency or accuracy, the correlation weakens, yielding a regression coefficient closer to 0.

Neutrality plots are biologically significant as they provide insights into the evolutionary forces shaping codon usage. A regression coefficient near 1 suggests that mutational constraints are largely responsible for variations in nucleotide composition, indicating that the genome is shaped by random drift or mutational equilibrium. On the other hand, a coefficient closer to 0 points to the dominance of natural or translational selection, suggesting that codon usage is being optimized for specific biological functions, such as increased translational efficiency or accuracy. The neutrality plot for all species across the Staphylococcus genus is presented in **Fig. 3A-3F**, illustrating the relative contributions of mutational pressure and natural selection on codon usage patterns.

Our results showed regression coefficients for CoPS (coagulase-positive staphylococci) and CoNS (coagulase-negative staphylococci) that were both close to 0, with values of -0.015 and -0.026, respectively. These negative coefficients imply that natural or translational selection, rather than mutational pressure, plays the dominant role in shaping codon usage within these genomes. This suggests that codons are being preferentially selected to optimize the translation process, likely enhancing the

organism's protein synthesis efficiency or accuracy. Similarly, for the co-variable species, the regression coefficient, while positive at 0.05, remains near 0, further supporting the conclusion that selective pressures rather than mutational bias are predominant in shaping codon usage.

These findings collectively indicate that, the neutrality plot analysis across CoPS, CoNS, and co-variable species reveals that mutational pressure has a minimal effect on codon usage bias, with natural or translational selection being the primary evolutionary force. This suggests that the genomes of these organisms are optimized for efficient translation, reflecting an adaptation that likely enhances their fitness by promoting effective protein synthesis under various environmental or physiological conditions.

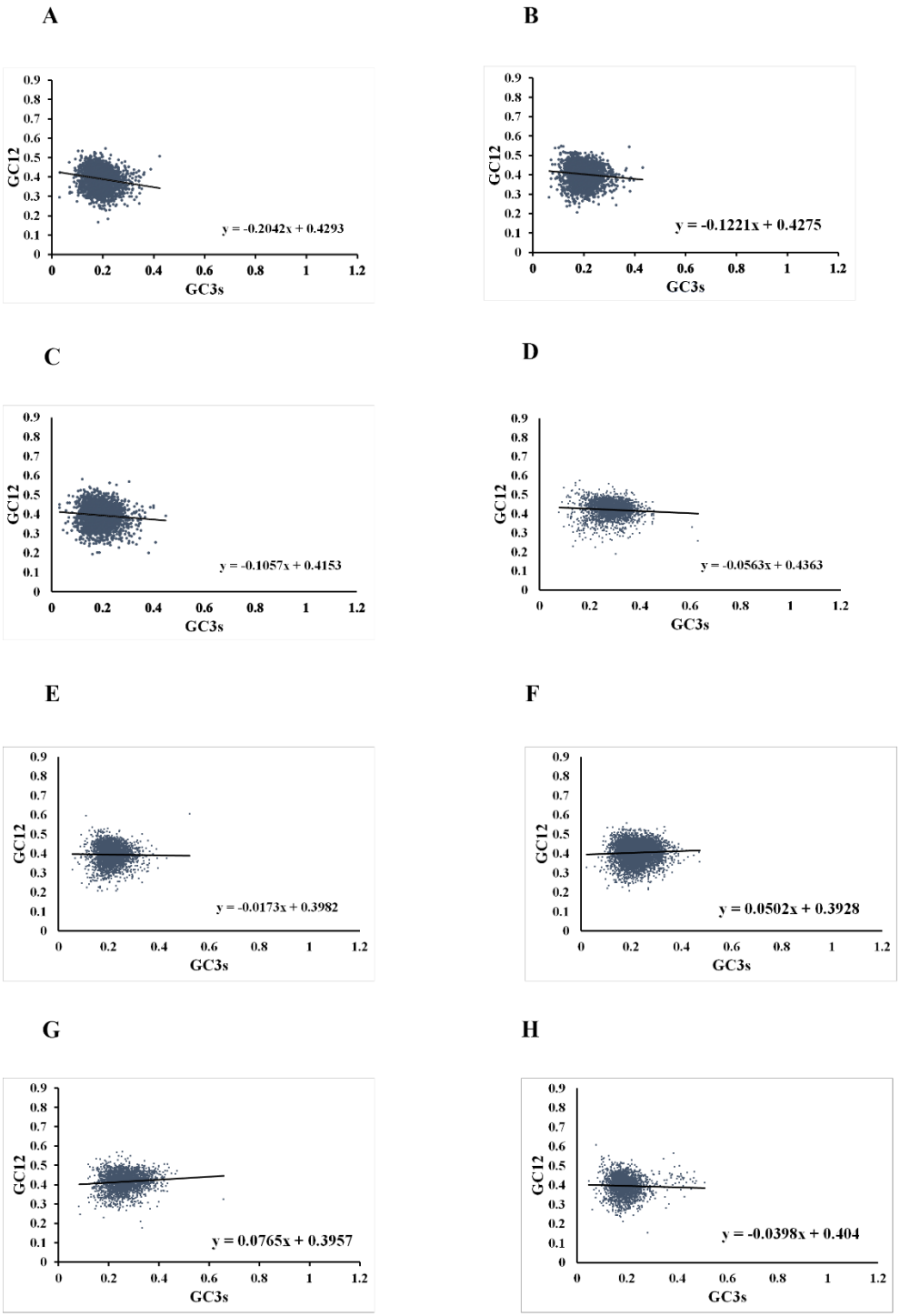


Figure 3A: GC3 v/s GC12 Plot with each data point representing an individual gene A). *Staphylococcus argenteus str. 58113* B). *Staphylococcus arlettae* C). *Staphylococcus auricularis* D). *Staphylococcus capitis* E). *Staphylococcus caprae M23864:W1* F). *Staphylococcus carnosus subsp. carnosus* G). *Staphylococcus chromogenes* H). *Staphylococcus cohnii*

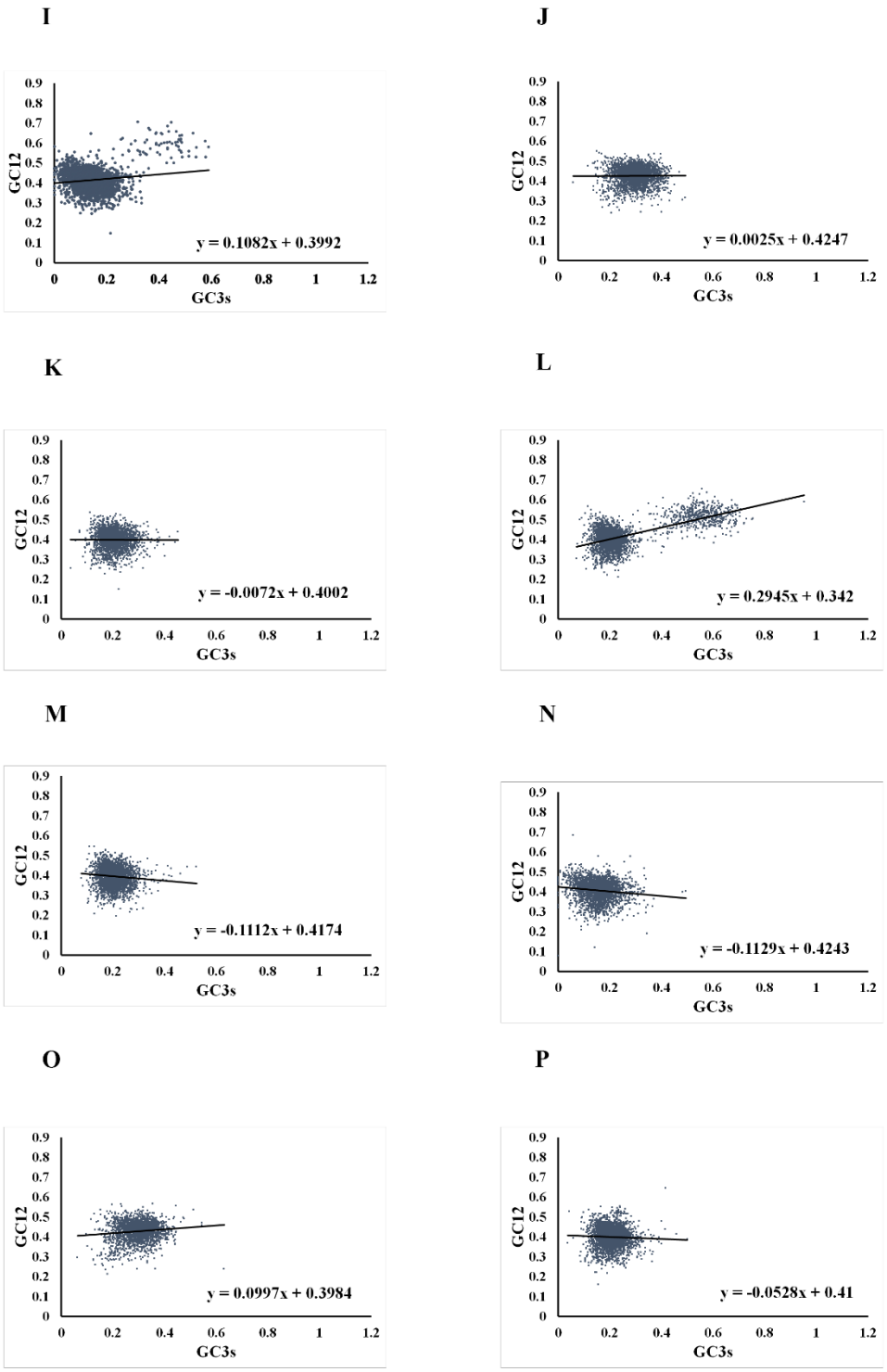


Figure 3B: GC3 v/s GC12 Plot with each data point representing an individual gene I). *Staphylococcus condimentistr. 58113* J). *Staphylococcus devriesei* K). *Staphylococcus edaphicus str. CCM 8730* L). *Staphylococcus equorum* M). *Staphylococcus felis M23864:W1* N). *Staphylococcus felis* O). *Staphylococcus fleurettii* P). *Staphylococcus gallinarum*

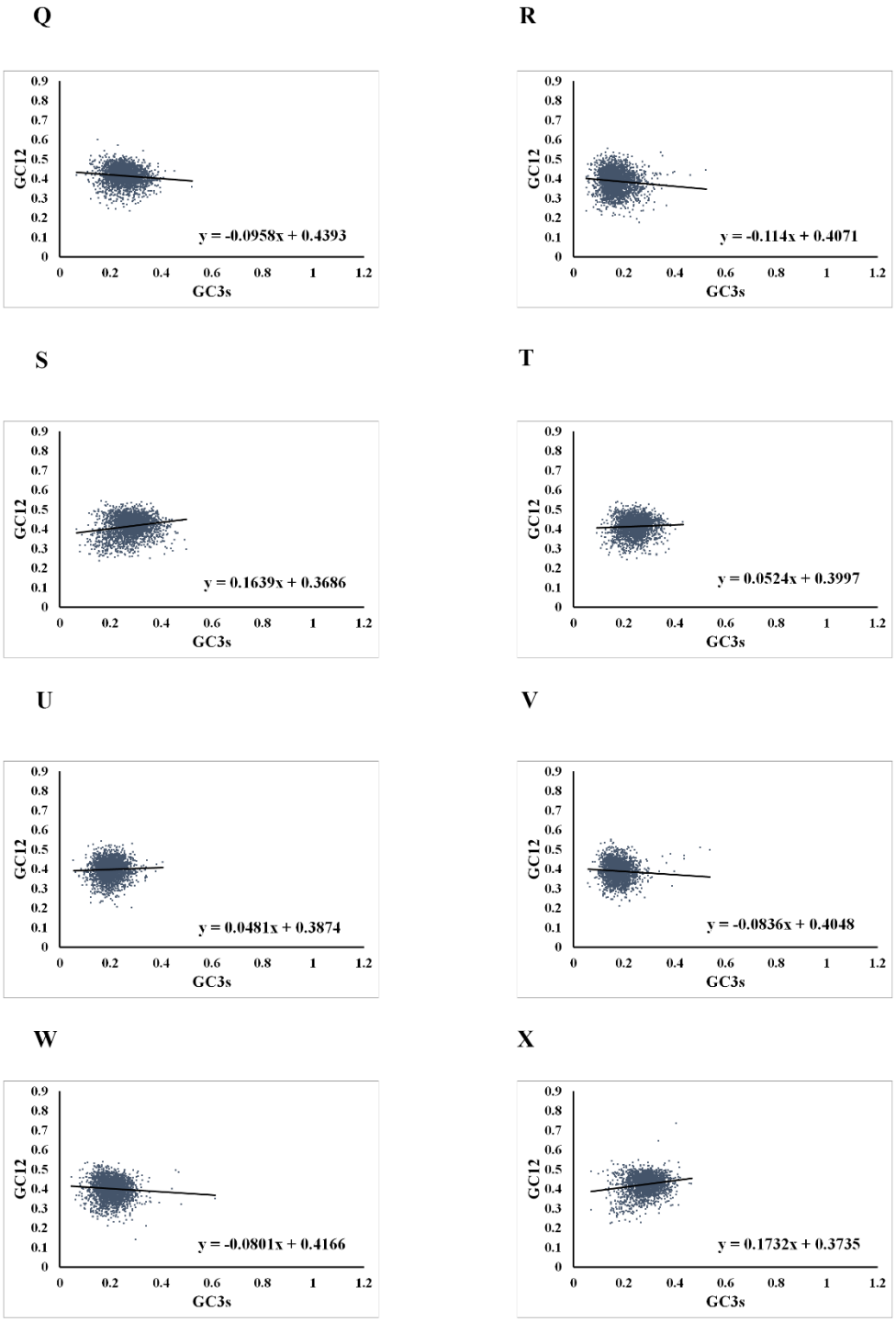


Figure 3C: GC3 v/s GC12 Plot with each data point representing an individual gene Q). *Staphylococcus haemolyticus* R). *Staphylococcus hominis* S). *Staphylococcus hyicus* T). *Staphylococcus intermedius NCTC 11048* U). *Staphylococcus kloosii* V). *Staphylococcus lentus* W). *Staphylococcus lugdunensis* X). *Staphylococcus lutrae*

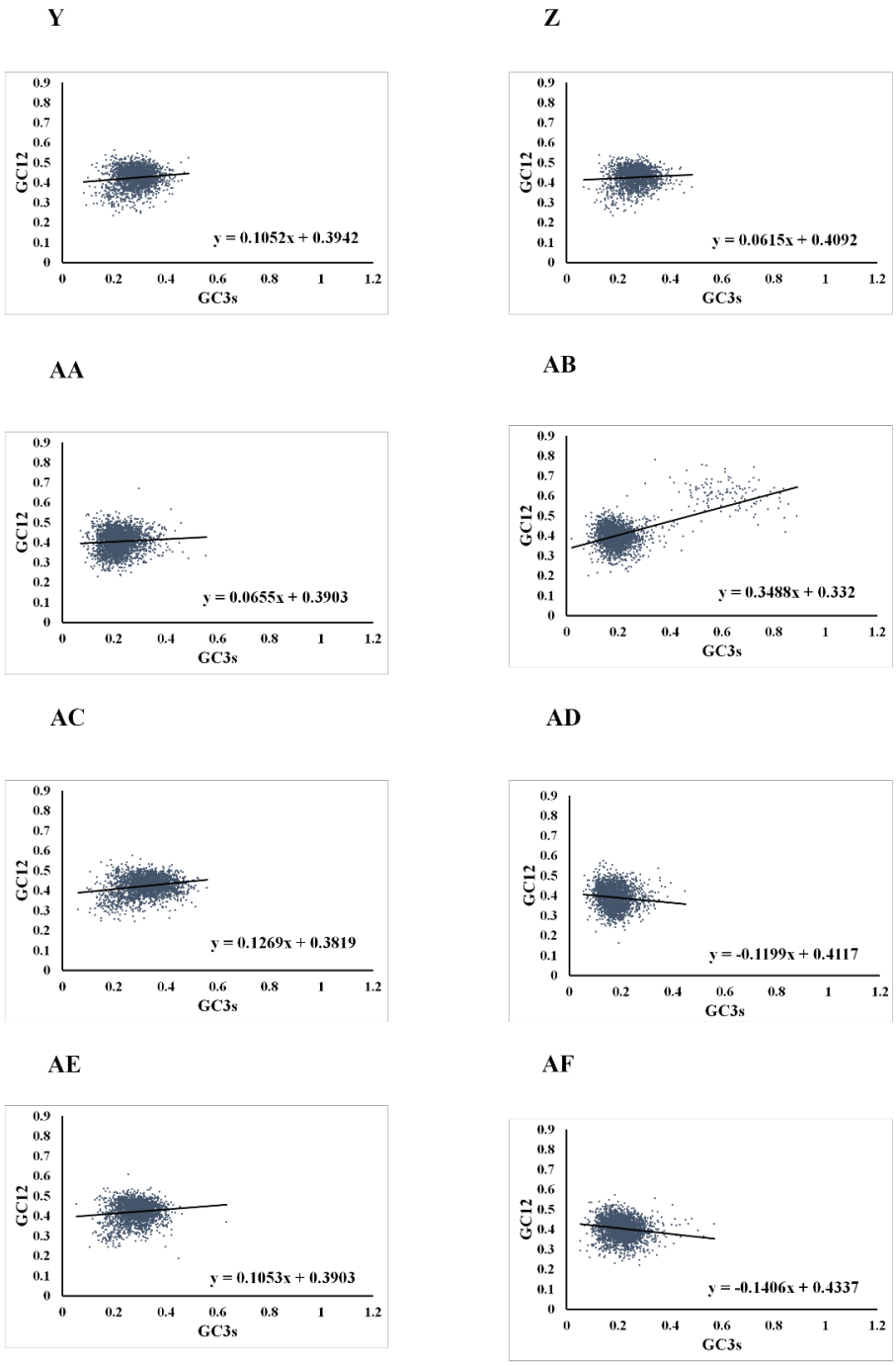


Figure 3D: GC3 v/s GC12 Plot with each data point representing an individual gene Y). *Staphylococcus massiliensis S46* Z). *Staphylococcus microti* AA). *Staphylococcus muscae* AB). *Staphylococcus nepalensis* AC). *Staphylococcus pasteuri* AD). *Staphylococcus petrasii subsp. croceilyticus* AE). *Staphylococcus pettenkoferi* AF). *Staphylococcus piscifermentans*

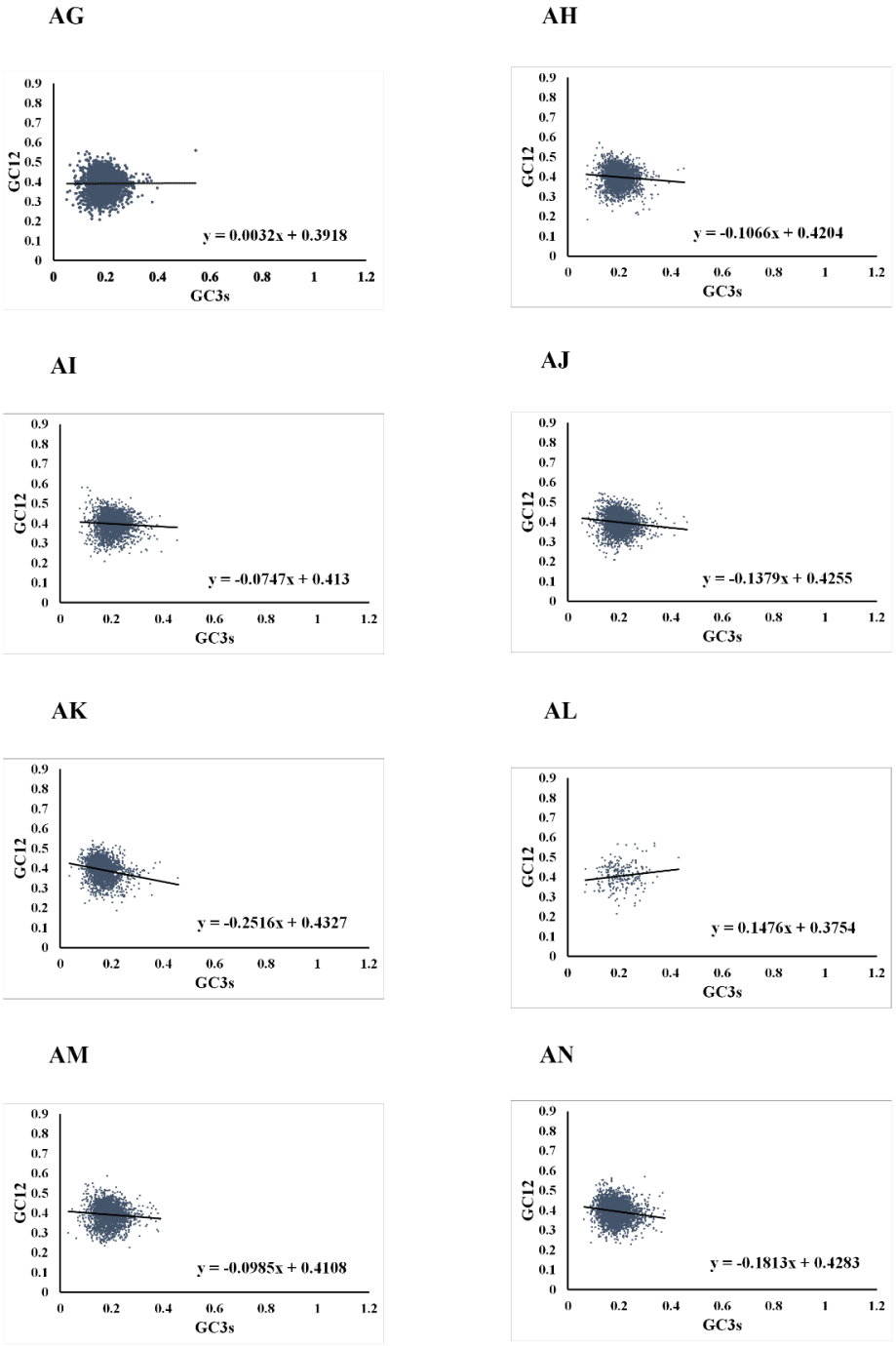


Figure 3E: GC3 v/s GC12 Plot with each data point representing an individual gene AG). *Staphylococcus pseudintermedius* AH). *Staphylococcus pseudoxylosus str. S04009* AI). *Staphylococcus rostri* AJ). *Staphylococcus saccharolyticus* AK). *Staphylococcus saprophyticus* AL). *Staphylococcus schleiferi* AM). *Staphylococcus sciuri* AN). *Staphylococcus simiae CCM 7213 = CCUG 51256*

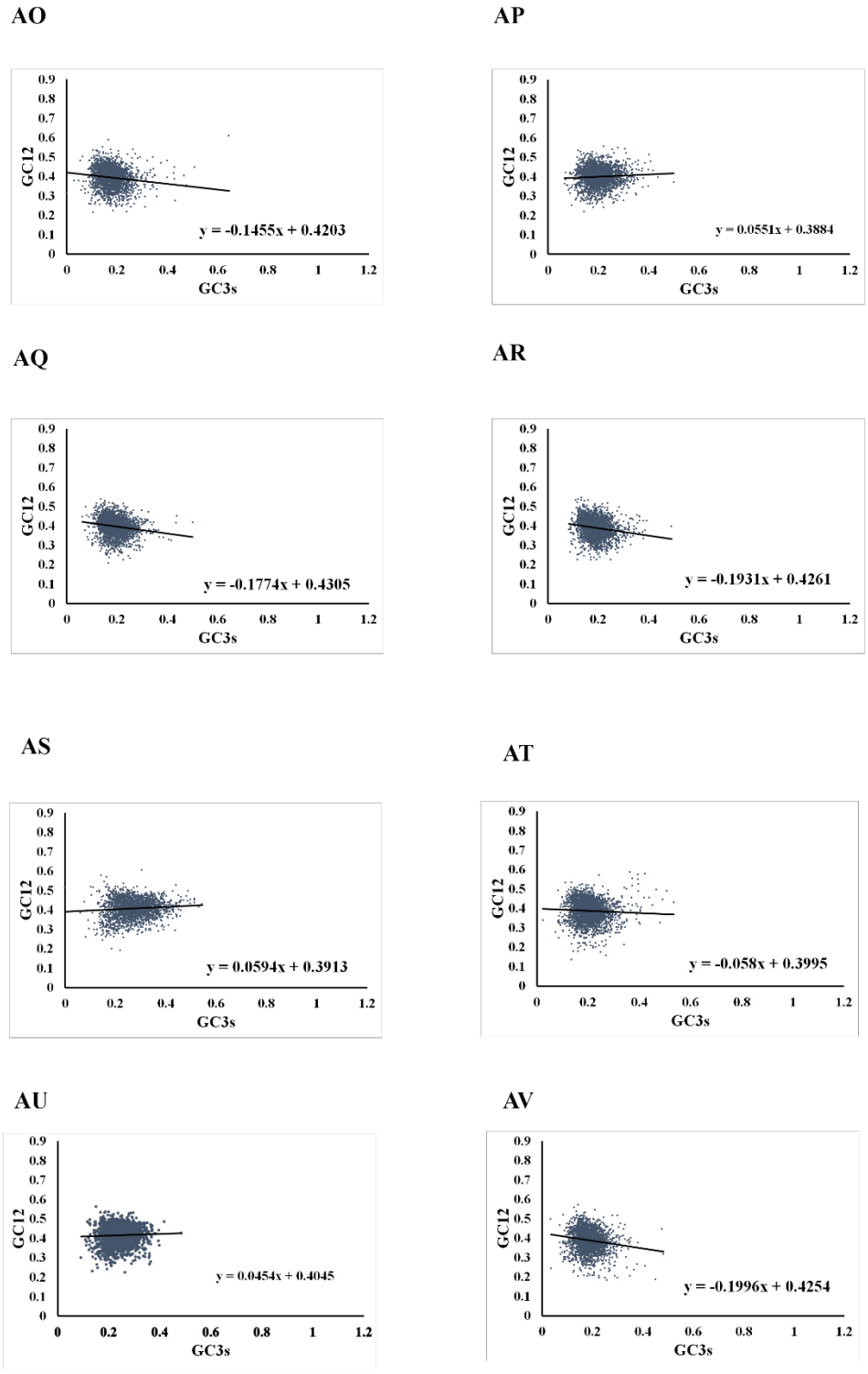


Figure 3F: GC3 v/s GC12 Plot with each data point representing an individual gene AO). *Staphylococcus simulans* AP). *Staphylococcus succinus* AQ). *Staphylococcus vitulinus* AR). *Staphylococcus warneri* AS). Staphylococcus xylosus AT) *Staphylococcus aureus* AU) *Staphylococcus agnetis* AV) *Staphylococcus epidermidis*

5.4. Preferential Purine Utilization in *Staphylococcus* Genomes

Using parity plot analysis (PR2), which enabled a more thorough investigation of nucleotide composition biases at the third codon position (GC3s and AT3s), we further investigated the bias in purine and pyrimidine usage within intrastent compositions in addition to the neutrality and ENc plot analysis. Plotting the ratio of G3s/(G3s + C3s), which represents GC bias, versus A3s/(A3s + T3s), which represents AT bias, was how the parity plot was specifically created. Chargaff's second parity rule, according to which the number of adenines (A) in a single strand of DNA is equal to the number of thymines (T) and the number of guanines (G) equals the number of cytosines (C), serves as the foundation for this analytical method. According to this rule, parity is achieved when both the GC bias and AT bias values reach 0.5, indicating an equal distribution of purines (A and G) and pyrimidines (T and C). The rationale for employing PR2 analysis lies in its ability to highlight deviations from this equilibrium, which may indicate evolutionary pressures such as mutational biases or selective forces acting on the nucleotide composition. Values above 0.5 on the plot indicate a preference for purines (A and G) over pyrimidines (T and C), suggesting that the evolutionary forces shaping codon usage are not solely driven by random mutational pressures but may also involve selective constraints. Conversely, values below 0.5 suggest a preference for pyrimidines, which would imply different mutational or selective forces at play.

According to our study, the mean AT and GC bias values continuously surpassed 0.5 for the whole Staphylococcus genus. This suggests that in protein-coding genes, adenine (A) is significantly preferred over thymine (T) and guanine (G) over cytosine (C) at the third codon position. Such a bias suggests that, within the genus, codon usage is influenced more by purine enrichment, potentially reflecting adaptive processes such as selection for translational efficiency or other genomic constraints favoring certain nucleotide compositions. The parity plot for all species across the Staphylococcus genus is shown in **Fig. 4A-4F**, illustrating these trends in purine and pyrimidine bias across different species. The observed purine dominance may also suggest an evolutionary advantage conferred by purine-rich codons, potentially related to the availability of tRNA pools or the energetic costs associated with nucleotide synthesis and usage during protein translation. Overall, these findings from the PR2 analysis

contribute to a deeper understanding of the codon usage patterns within the Staphylococcus genus, highlighting the non-random nature of nucleotide composition at synonymous sites and the role of selective pressures in shaping genome evolution.

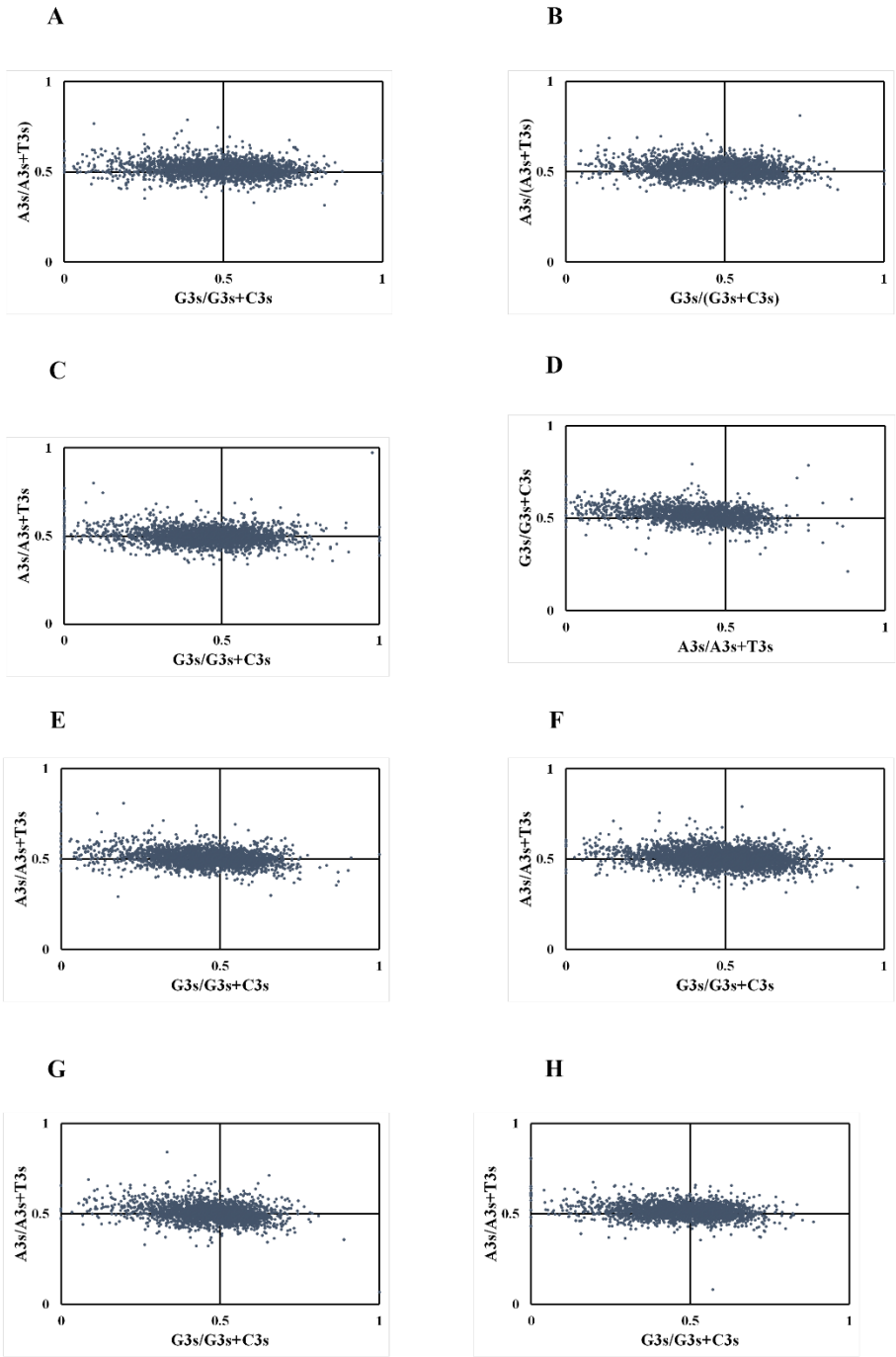


Figure 4A: G3s/ G3s + C3s Vs A3s/ A3s + T3s plot with each data point representing an individual gene A). *Staphylococcus argenteus str. 58113* B). *Staphylococcus arlettae* C). *Staphylococcus auricularis* D). *Staphylococcus capitis* E). *Staphylococcus caprae M23864:W1* F). *Staphylococcus carnosus subsp. carnosus* G). *Staphylococcus chromogenes* H). *Staphylococcus cohnii*

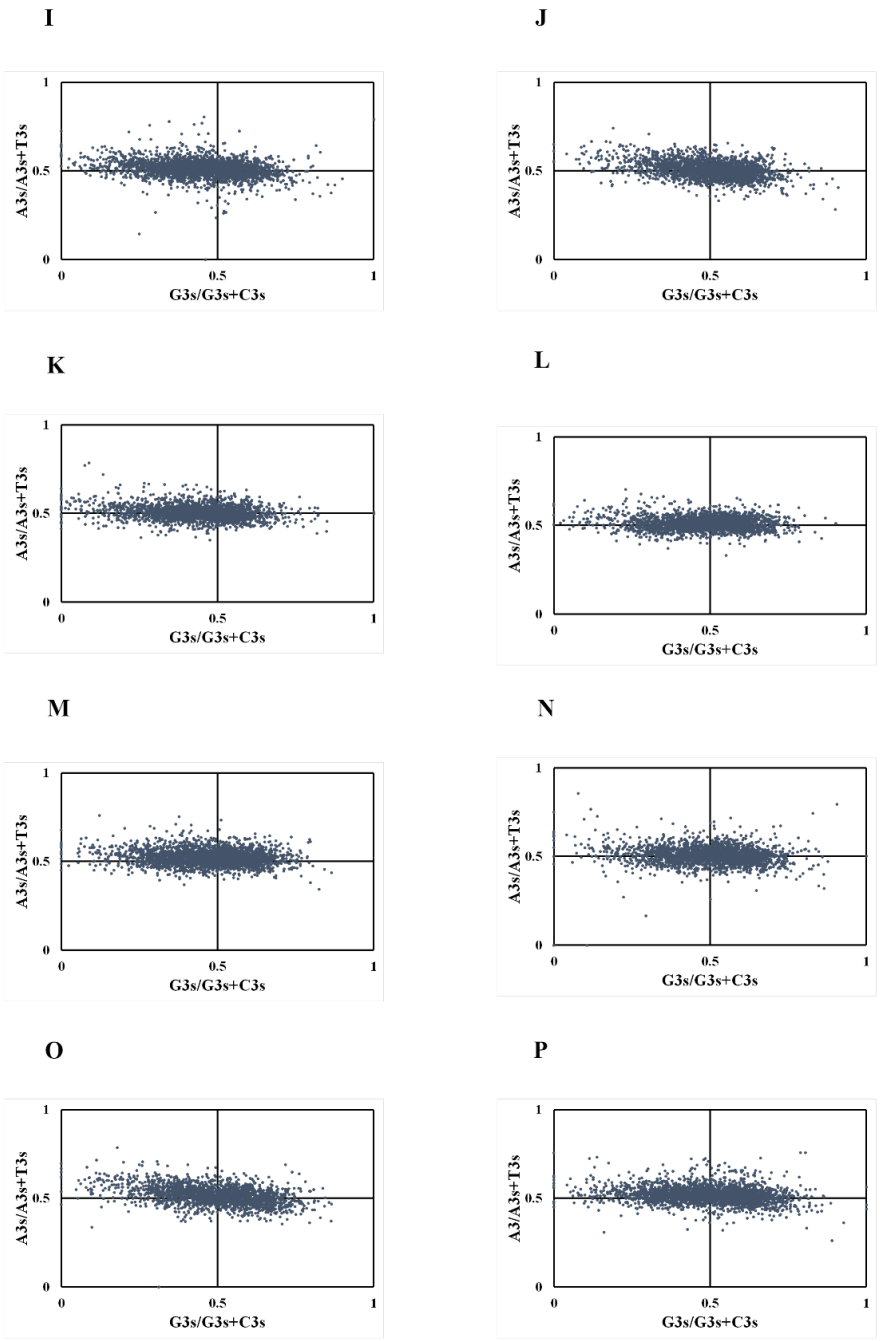


Figure 4B: G3s/ G3s + C3s Vs A3s/ A3s + T3s plot with each data point representing an individual gene I). *Staphylococcus condimentistr. 58113* J). *Staphylococcus devriesei* K). *Staphylococcus edaphicus str. CCM 8730* L). *Staphylococcus equorum* M). *Staphylococcus felis M23864:W1* N). *Staphylococcus felis* O). *Staphylococcus fleurettii* P). *Staphylococcus gallinarum*

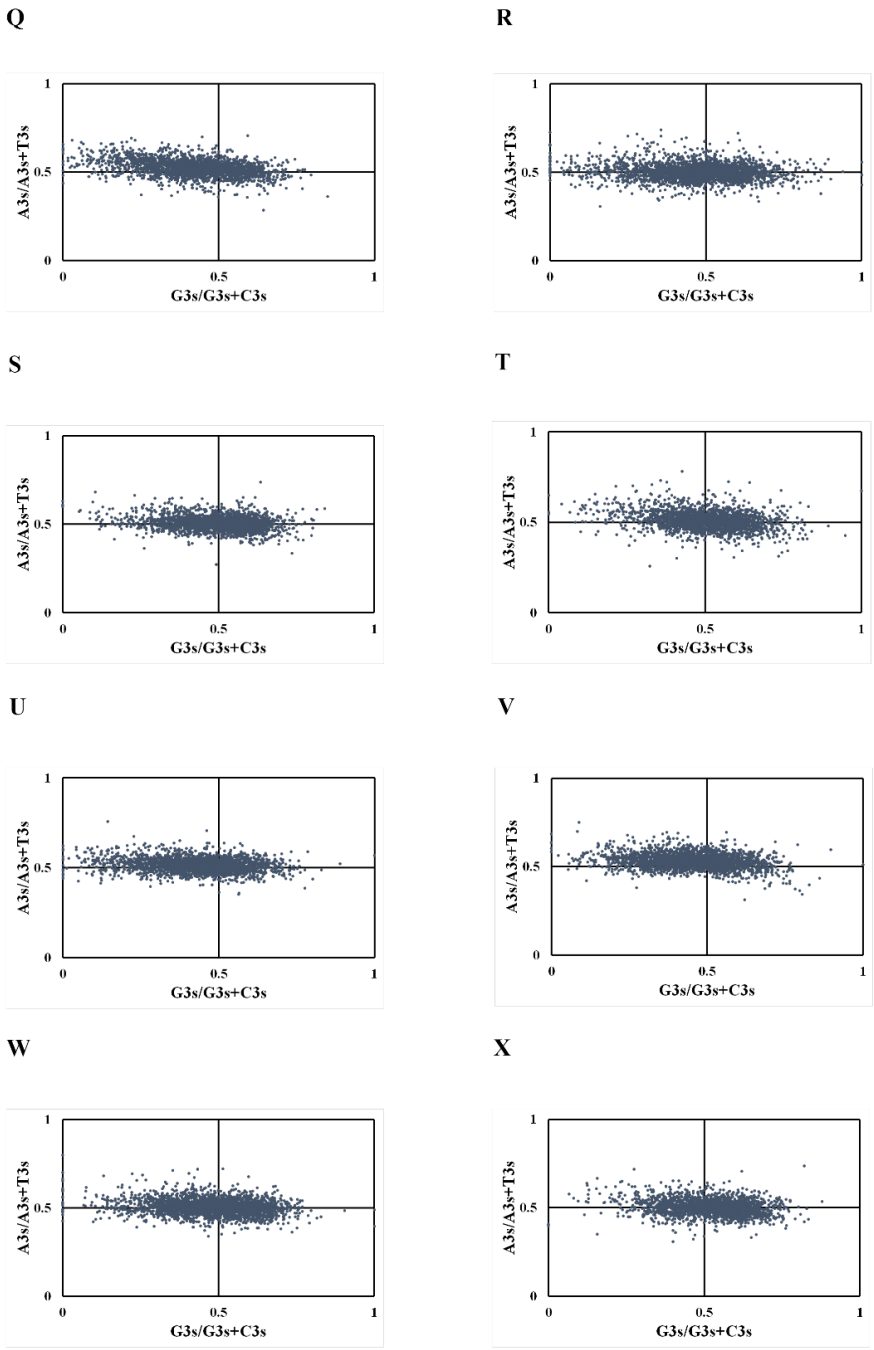


Figure 4C: G3s/ G3s + C3s Vs A3s/ A3s + T3s plot with each data point representing an individual gene Q). *Staphylococcus haemolyticus* R). *Staphylococcus hominis* S). *Staphylococcus hyicus* T). *Staphylococcus intermedius NCTC 11048* U). *Staphylococcus kloosii* V). *Staphylococcus lentus* W). *Staphylococcus lugdunensis* X). *Staphylococcus lutrae*

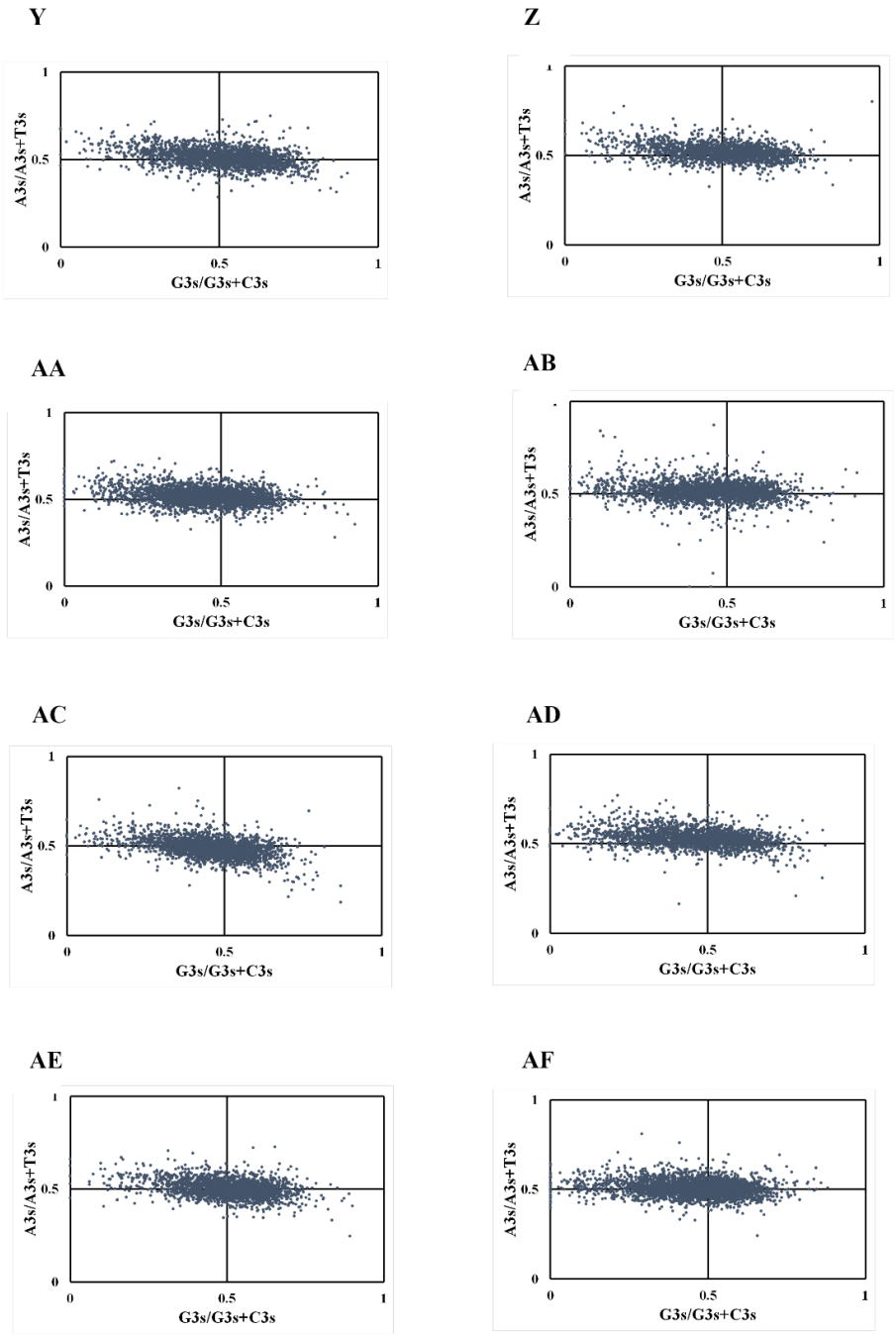


Figure 4D: G3s/ G3s + C3s Vs A3s/ A3s + T3s plot with each data point representing an individual gene Y). *Staphylococcus massiliensis S46* Z). *Staphylococcus microti* AA). *Staphylococcus muscae* AB). *Staphylococcus nepalensis* AC). *Staphylococcus pasteuri* AD). *Staphylococcus petrasii subsp. croceilyticus* AE). *Staphylococcus pettenkoferi* AF). *Staphylococcus piscifermentans*

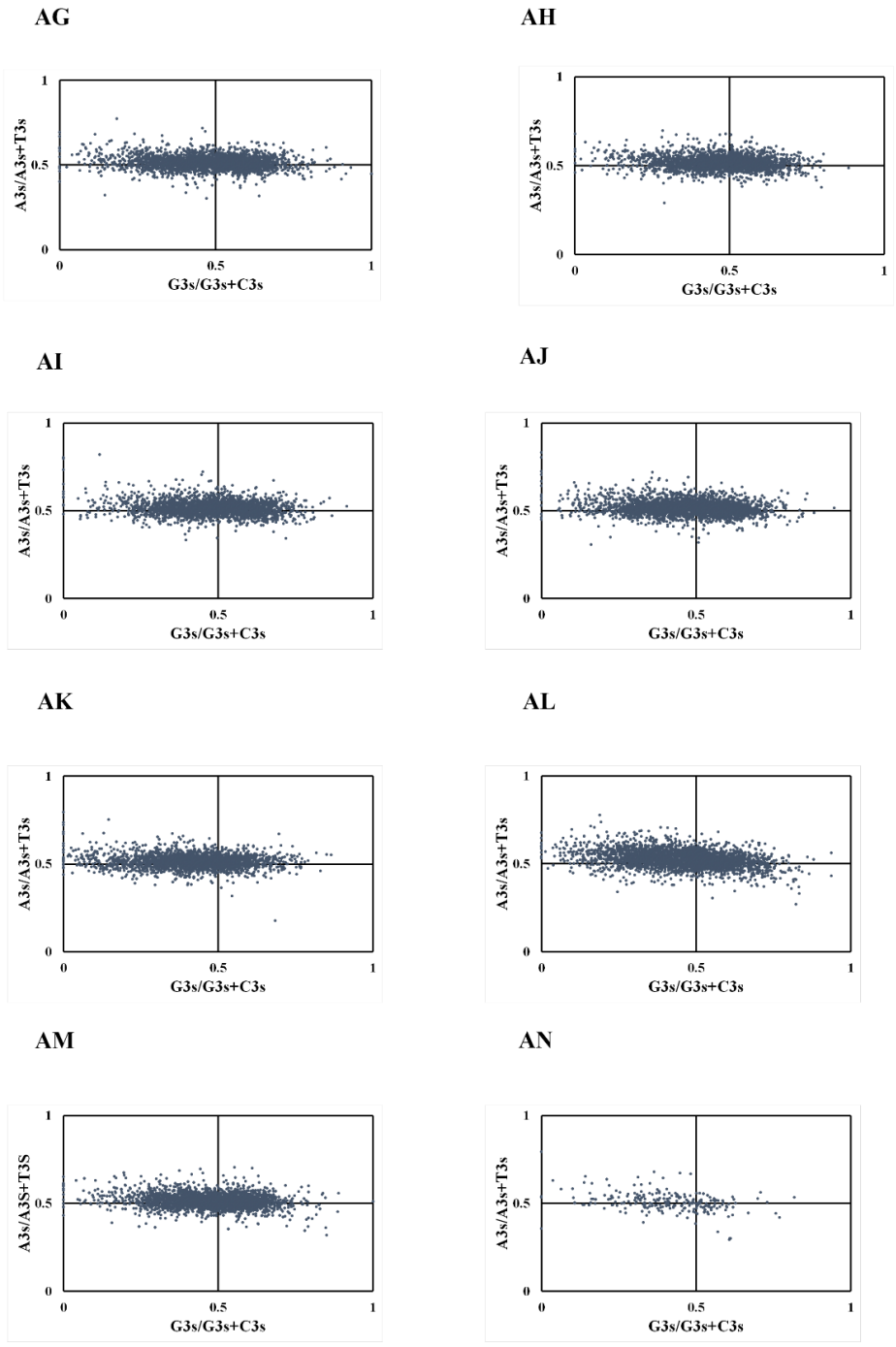


Figure 4E: G3s/ G3s + C3s Vs A3s/ A3s + T3s plot with each data point representing an individual gene AG). *Staphylococcus pseudintermedius* AH). *Staphylococcus pseudoxylosus str. S04009* AI). *Staphylococcus rostri* AJ). *Staphylococcus saccharolyticus* AK). *Staphylococcus saprophyticus* AL). *Staphylococcus schleiferi* AM). *Staphylococcus sciuri* AN). *Staphylococcus simiae CCM 7213 = CCUG 51256*

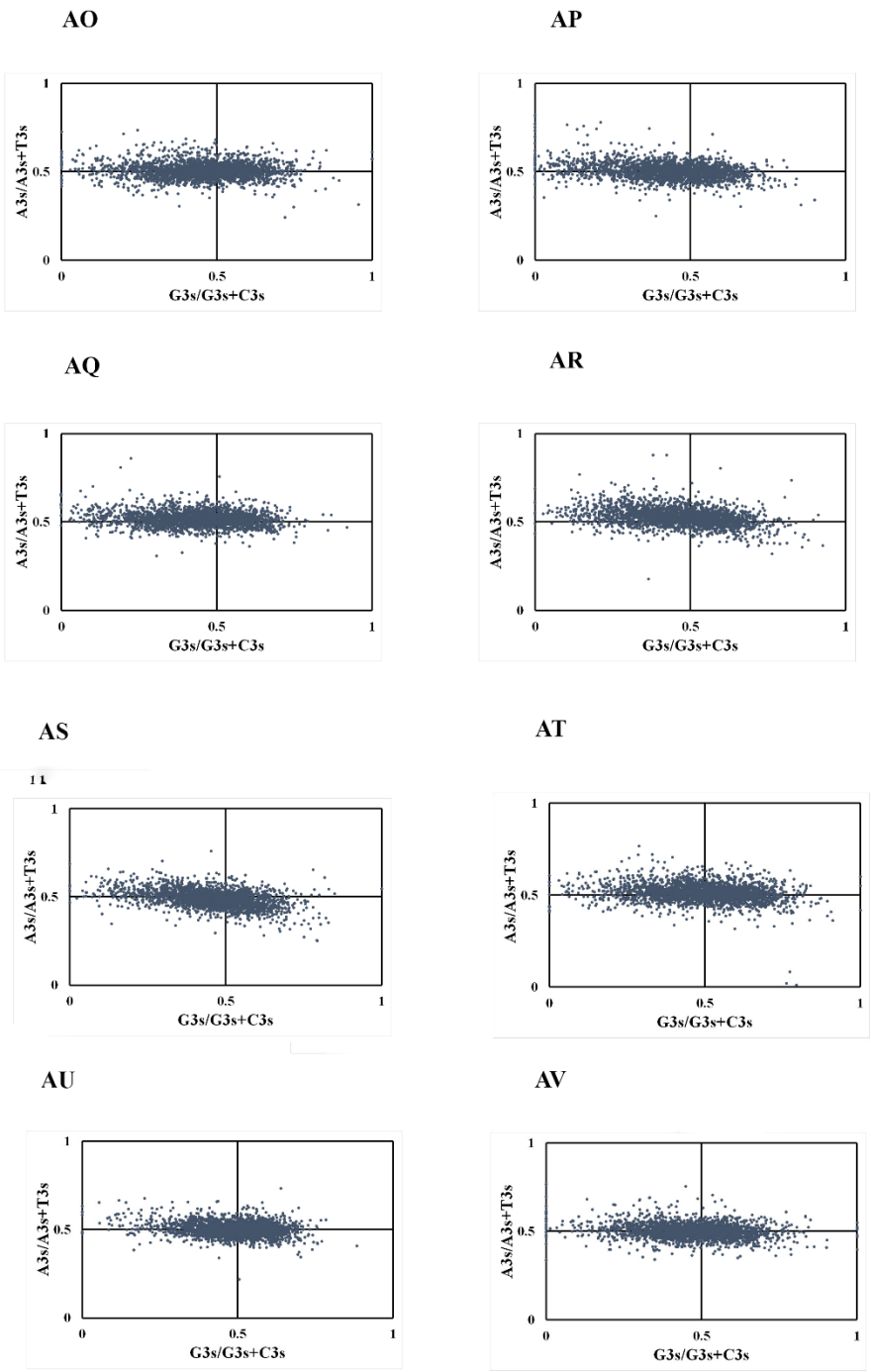


Figure 4F: G3s/ G3s + C3s Vs A3s/ A3s + T3s plot with each data point representing an individual gene AO). *Staphylococcus simulans* AP). *Staphylococcus succinus* AQ). *Staphylococcus vitulinus* AR). *Staphylococcus warneri* AS). *Staphylococcus xylosus* AT) *Staphylococcus aureus* AU) *Staphylococcus agnetis* AV) *Staphylococcus epidermidis*

5.5. Distinct Dinucleotide Patterns in *Staphylococcus* Variants

The host immune system relies on pattern recognition receptors (PRRs) to detect pathogen-associated molecular patterns (PAMPs), which are molecular markers characteristic of microbial pathogens. These markers include dinucleotide pairs combinations of two nucleotides that function as PAMPs and stimulate immune responses, as recent research indicates [165]. Among the four nucleotides adenine (A), thymine (T), guanine (G), and cytosine (C) various combinations yield 16 possible dinucleotide pairs, each playing unique roles in microbial detection and immune system interactions. One pair, CpG (cytosine-phosphate-guanine), is particularly significant for its immunostimulatory potential, as it is recognized by Toll-like Receptor 9 (TLR-9). TLR-9 identifies bacterial and viral DNA containing CpG motifs, triggering essential immune responses that mobilize immune cells and initiate inflammatory pathways [166]. However, some pathogens can strategically lower the presence of CpG motifs within their DNA to avoid TLR-9 detection, thereby evading immune surveillance and enhancing pathogenicity by reducing recognition likelihood.

Our analysis of dinucleotide frequency in various Staphylococcus species—including coagulase-positive staphylococci (CoPS), coagulase-negative staphylococci (CoNS), and co-variable species (CoVS) revealed distinctive patterns, suggesting these dinucleotides play significant roles in genome structure and immune evasion strategies. Specifically, GpT was found to be particularly prevalent in CoPS species, while GpC was abundant in CoNS species, and CpA was prominent in CoVS. These observed enrichments imply that each dinucleotide may confer adaptive advantages within these groups. Conversely, certain dinucleotides showed consistently low prevalence across species; for instance, CpT was minimal in both CoPS and CoVS, and TpA was the least frequent in CoNS (**Table 2a-2c**). Notably, CpG was universally scarce across CoPS, CoNS, and CoVS, supporting the hypothesis that pathogens might downregulate CpG content to avoid immune detection.

The reduced CpG presence likely reflects an evolutionary adaptation that aids these species in bypassing immune surveillance, particularly TLR-9 recognition. Since TLR-9 is critical for detecting CpG motifs and initiating a cascade of immune responses to counter bacterial and viral invasions, selectively lowering CpG content enables

pathogens to reduce their immune detection footprint and enhance survival. This evasion strategy may contribute significantly to the persistence of Staphylococcus infections, including conditions that involve immune modulation, such as certain sexually transmitted infections (STIs) and chronic bacterial infections that suppress immune responses [167]. Overall, these findings indicate that dinucleotide composition, especially reduced CpG motifs, is central to immune evasion in CoPS, CoNS, and CoVS species. Such dinucleotide biases not only impact immune detection avoidance but also influence pathogenicity, providing valuable insights into the selective pressures shaping microbial genomes and the evolutionary mechanisms that support bacterial survival and persistent infection (**Fig. 5.**).

Table 2a: Relative dinucleotide abundance analysis for CoPS

Species	Dinucleotide															
	AA	**AC**	**AG**	**AT**	**CA**	**CC**	**CG**	**CT**	**GA**	**GC**	**GG**	**GT**	**TA**	**TC**	**TG**	**TT**
Staphylococcus argenteus	1.049898	0.990078	0.89418	1.001491	1.247871	0.861715	0.941074	0.804597	0.970671	1.280693	0.943644	0.934563	0.83299	0.904131	1.187327	1.129829
Staphylococcus aureus	1.047703	0.987194	0.89431	1.005885	1.239816	0.852644	0.958326	0.810032	0.975814	1.27961	0.936341	5.305388	0.931483	3.292691	0.8347	0.912224
Staphylococcus delphini	1.095329	1.030616	0.797967	1.014056	1.231837	0.776318	1.248526	0.685315	1.016631	1.161031	0.871242	0.974796	0.723151	0.982531	1.173501	1.198013
Staphylococcus intermedius	1.087222	1.018801	0.816938	1.010763	1.237869	0.794885	1.179561	0.713292	1.014668	1.164622	0.888241	0.96331	0.735422	0.97974	1.187276	1.183804

Staphylococcus lutrae	1.095832	0.981389	0.796448	1.043038	1.242731	0.847948	1.150097	0.705968	1.028892	1.17151	0.946045	0.90147	0.715338	0.985645	1.179001	1.198657
Staphylococcus pseudintermedius	1.092861	1.023639	0.809252	1.00914	1.229093	0.799766	1.203992	0.7096	1.011307	1.170407	0.892994	0.960816	0.74018	0.967915	1.173013	1.189221

Table 2b: Relative dinucleotide abundance analysis for CoVS

	Dinucleotide															
Species	**AA**	**AC**	**AG**	**AT**	**CA**	**CC**	**CG**	**CT**	**GA**	**GC**	**GG**	**GT**	**TA**	**TC**	**TG**	**TT**
Staphylococcus agnetis	1.082038	1.01267	0.804056	1.020813	1.210898	0.853936	1.122476	0.759683	0.968505	1.200407	0.966481	0.944661	0.801248	0.933003	1.175872	1.145776
Staphylococcus hyicus	1.087742	1.00594	0.801303	1.019491	1.209435	0.864917	1.125832	0.749915	0.969591	1.202049	0.973205	0.938738	0.794239	0.932516	1.175364	1.156088

Table 2c: Relative dinucleotide abundance analysis for CoNS

	Dinucleotide															
Species	**AA**	**AC**	**AG**	**AT**	**CA**	**CC**	**CG**	**CT**	**GA**	**GC**	**GG**	**GT**	**TA**	**TC**	**TG**	**TT**
Staphylococcus arlettae	1.027994	1.038898	0.917575	0.990492	1.229298	0.882727	0.910668	0.84703	0.937611	1.283491	0.957922	0.955517	0.88337	0.835532	1.165493	1.114104

Staphylococcus auricularis	1.052458	1.014162	0.852079	1.024029	1.171074	0.764798	1.154913	0.834251	1.047534	1.082531	0.889097	0.967456	0.783083	1.074426	1.156597	1.099938
Staphylococcus capitis	1.044017	0.973049	0.946836	0.984731	1.140009	0.879043	0.883835	0.965929	1.01546	1.159902	0.974526	0.91674	0.862498	0.992301	1.13124	1.083559
Staphylococcus caprae	1.051477	0.97111	0.922343	0.993301	1.153084	0.856158	0.964811	0.913115	1.026081	1.152361	0.956357	0.915819	0.83539	1.011695	1.135977	1.104428
Staphylococcus carnosus	1.055971	0.970657	0.923611	0.987483	1.133147	0.861588	0.997884	0.914951	0.985506	1.236093	0.950075	0.92768	0.868702	0.957364	1.119415	1.100455
Staphylococcus chromogenes	1.110022	0.981179	0.824248	0.995484	1.156654	0.908602	1.110186	0.796254	1.010261	1.160977	0.987971	0.899737	0.764259	0.96385	1.145822	1.192685
Staphylococcus cohnii	1.047463	0.98649	0.905136	1.000078	1.236346	0.907955	0.870679	0.845312	0.968845	1.279775	0.974719	0.916048	0.841487	0.887213	1.188088	1.124543
Staphylococcus condimenti	1.084684	0.931405	0.897843	0.995729	1.179958	0.861026	0.983019	0.87249	1.008507	1.288254	0.913803	0.888912	0.786579	0.962704	1.186919	1.148074
Staphylococcus devriesei	1.051596	0.991715	0.925812	0.980873	1.081513	0.891217	1.03036	0.940129	0.983654	1.217792	0.963333	0.928868	0.899171	0.92871	1.093389	1.096385

Staphylococcus edaphicus	1.03797	0.979145	0.902579	1.018513	1.23085	0.910268	0.904394	0.832868	0.969409	1.26831	0.975735	0.91577	0.851429	0.898284	1.176438	1.113321
Staphylococcus epidermidis	1.043692	0.981812	0.926926	0.991788	1.186163	0.880289	0.853567	0.923859	1.006136	1.170173	0.991077	0.914316	0.847904	0.976388	1.158848	1.095961
Staphylococcus equorum	1.038309	0.999347	0.925111	0.992731	1.199291	0.879125	0.92066	0.869855	0.983448	1.255236	0.951161	0.922507	0.857376	0.898078	1.158887	1.120724
Staphylococcus felis	1.054899	0.965613	0.886655	1.018261	1.223952	0.854406	0.989545	0.824266	1.013453	1.212306	0.953687	0.900597	0.801519	0.976879	1.164028	1.134225
Staphylococcus fleurettii	1.025923	0.979753	0.971217	0.985876	1.192802	0.897056	0.795877	0.932448	1.01801	1.200543	0.959179	0.909951	0.862625	0.94925	1.149537	1.100707
Staphylococcus gallinarum	1.021691	0.994871	0.919374	1.019403	1.240792	0.899038	0.878665	0.843667	0.978215	1.236241	0.977245	0.922075	0.86228	0.904912	1.166808	1.101828
Staphylococcus haemolyticus	1.038304	0.99918	0.891422	1.015025	1.194082	0.849347	0.981931	0.859556	0.990486	1.202956	0.941367	0.941392	0.852578	0.947918	1.171745	1.090011
Staphylococcus hominis	1.048219	0.971533	0.926304	0.989544	1.169324	0.882296	0.871412	0.927529	1.009389	1.192468	0.976412	0.912484	0.852772	0.974851	1.156067	1.094148

Staphylococcus kloosii	1.044558	1.024171	0.908066	0.983248	1.150017	0.875561	1.037332	0.861845	0.932705	1.296605	0.973722	0.947491	0.906614	0.850771	1.104704	1.118699
Staphylococcus lentus	1.028892	0.982717	0.973579	0.980255	1.149939	0.912272	0.86691	0.937669	1.011843	1.211039	0.974266	0.902907	0.881092	0.932354	1.107291	1.110634
Staphylococcus lugdunensis	1.042595	0.976706	0.892673	1.020069	1.247446	0.873248	0.913864	0.831649	0.955839	1.318962	0.976418	0.901519	0.844556	0.893171	1.179954	1.121683
Staphylococcus massiliensis	1.070113	1.029273	0.888126	0.969476	1.057689	0.78318	1.230226	0.900492	1.021887	1.13966	0.885553	0.969964	0.861198	0.995457	1.073333	1.112055
Staphylococcus microti	1.061392	1.07951	0.830467	0.996811	1.248851	0.732383	1.196244	0.725258	0.995074	1.213585	0.840977	0.990758	0.768369	0.910814	1.196131	1.179026
Staphylococcus muscae	1.040385	1.057848	0.846589	1.020585	1.306164	0.737644	1.046877	0.764735	1.03054	1.137252	0.86456	0.978224	0.737564	0.988942	1.248831	1.131692
Staphylococcus nepalensis	1.053712	0.976442	0.905794	0.99944	1.189368	0.928271	0.962968	0.836434	0.956921	1.306385	0.980284	0.905045	0.859201	0.8676	1.140049	1.142915
Staphylococcus pasteuri	1.026794	0.978897	0.922473	1.017337	1.209489	0.906552	0.897006	0.860257	0.994344	1.165522	0.974603	0.937321	0.858842	0.969089	1.157461	1.086163
Staphylococcus petrasii	1.042083	0.973487	0.953044	0.985909	1.142702	0.821548	0.995429	0.925937	1.007324	1.24386	0.922339	0.909407	0.862199	0.970389	1.105562	1.111986

Staphylococcus pettenkoferi	1.027251	1.02051	0.928867	0.998829	1.106674	0.690336	1.189352	0.937384	1.095857	1.093973	0.81957	0.96073	0.813572	1.110148	1.093136	1.072032
Staphylococcus piscifermentans	1.039786	0.995913	0.919658	0.996803	1.208497	0.878817	0.919116	0.864123	0.976796	1.269406	0.950136	0.92384	0.856693	0.894809	1.164489	1.117177
Staphylococcus pseudoxylosus	1.037415	0.996956	0.925151	0.9957	1.200473	0.925321	0.883228	0.873897	0.965754	1.267318	0.981197	0.919475	0.871216	0.871595	1.155436	1.115909
Staphylococcus rostri	1.05619	1.083828	0.820396	1.009236	1.255793	0.725618	1.199308	0.723215	0.982812	1.20202	0.852373	1.00226	0.775935	0.920976	1.194798	1.161288
Staphylococcus saccharolyticus	1.031289	0.992547	0.935491	0.995108	1.198615	0.900936	0.837272	0.90221	0.99033	1.184342	0.98003	0.934961	0.867903	0.947878	1.162925	1.087896
Staphylococcus saprophyticus	1.047843	0.982545	0.902407	1.002324	1.23306	0.911043	0.868231	0.848037	0.971381	1.277117	0.979587	0.9119	0.841609	0.893551	1.188749	1.121908
Staphylococcus schleiferi	1.06972	0.97657	0.872809	1.004026	1.227917	0.851281	1.018382	0.799376	0.985832	1.255783	0.936387	0.922423	0.800464	0.938298	1.178649	1.148611
Staphylococcus sciuri	1.020365	0.984635	0.945943	1.007155	1.222349	0.85216	0.899627	0.862103	1.022091	1.179857	0.935323	0.925351	0.848368	0.976904	1.150875	1.101064

Staphylococcus simiae	1.015424	1.02094	0.928357	1.005533	1.263217	0.875926	0.818455	0.85803	0.963331	1.248742	0.950516	0.952645	0.871933	0.886617	1.195755	1.08678
Staphylococcus simulans	1.06628	0.998194	0.901879	0.979955	1.20781	0.782691	1.067116	0.830183	1.003741	1.192248	0.87594	0.966874	0.78545	0.99588	1.164456	1.146575
Staphylococcus succinus	1.048489	0.956417	0.901764	1.021748	1.167706	0.944872	0.9439	0.870849	0.961377	1.30182	0.984155	0.895569	0.873132	0.879029	1.151184	1.109477
Staphylococcus vitulinus	1.030936	0.967116	0.956295	0.996062	1.178514	0.86691	0.909944	0.900553	1.023213	1.216851	0.938841	0.905931	0.856634	0.968244	1.131583	1.107783
Staphylococcus warneri	1.020704	0.987946	0.924841	1.018459	1.2282	0.899276	0.880485	0.84901	0.994132	1.15901	0.967776	0.945706	0.857395	0.966417	1.167879	1.084675
Staphylococcus xylosus	1.037526	0.989144	0.929576	0.995729	1.189067	0.929191	0.881742	0.883343	0.974887	1.263443	0.975336	0.916932	0.872646	0.883462	1.153517	1.110822

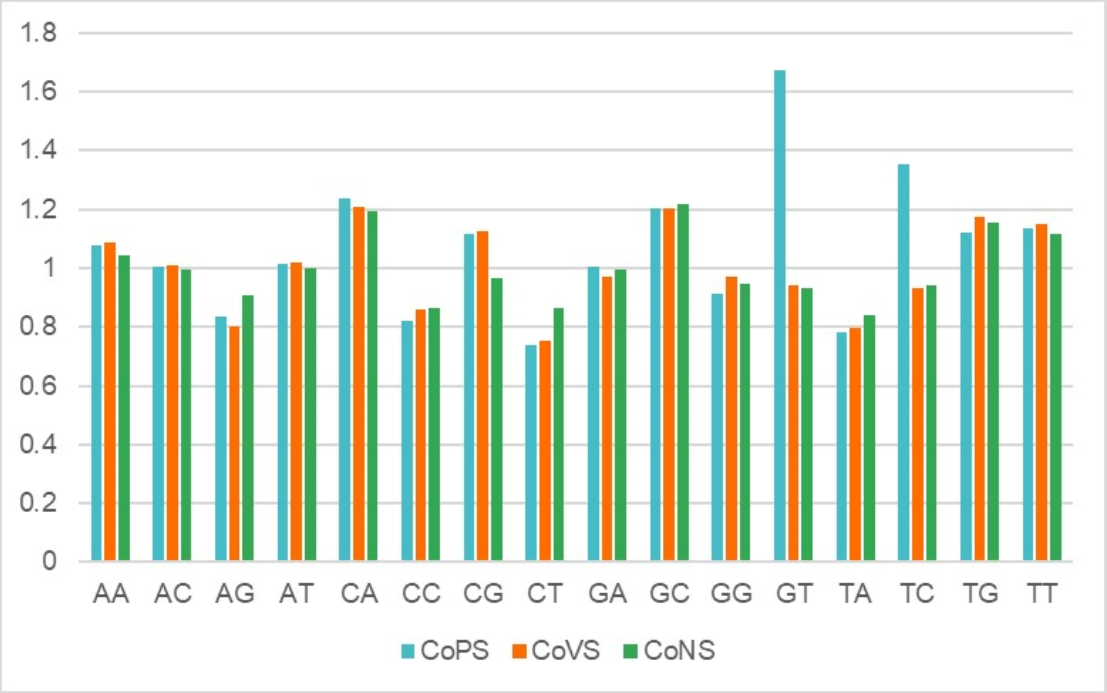


Fig. 5. Dinucleotide frequencies of CoPS, CoNS & CoVS

5.6. Codon Preferences in the Genomic Landscape

Codon usage bias is a fundamental aspect of all genomes, playing a crucial role in regulating gene expression, particularly through its impact on the process of translation [74]. This bias arises when certain synonymous codons (different codons that code for the same amino acid) are preferred over others. Understanding these preferences offers valuable insights into the evolutionary pressures and functional constraints that shape microbial genomes. In organisms like *Staphylococcus*, this bias helps optimize the translation process, ensuring that proteins essential for growth, adaptation, and virulence are synthesized efficiently. By favoring specific codons, these microbes can fine-tune gene expression, giving them an edge in survival, particularly in host environments.

In our analysis of codon usage patterns across the *Staphylococcus* genus, we observed strong codon bias in coagulase-positive staphylococci (CoPS), coagulase-negative staphylococci (CoNS), and co-variable species. One codon, TTA, which encodes leucine, emerged as the most frequently used codon across the genus. For CoPS, other preferred codons included CGT (arginine), GGT (glycine), and AGA (arginine),

highlighting their importance in genes crucial for microbial growth and virulence. In co-variable species, CGT (arginine) and ACA (threonine) were also favored, while in CoNS, CGT (arginine) and GGT (glycine) were prominent. On the other hand, codons like CCC (proline), TCC (serine), and CTG (leucine) were among the least used in CoPS, while CTG, CAG (glutamine), and CGG (arginine) were least preferred in co-variable species and CoNS (**Table 3a-3c**).

Our findings, based on relative synonymous codon usage (RSCU), point to a clear trend: AT-rich codons, especially TTA, are prevalent across the genus. This aligns with the idea that Staphylococcus bacteria may favor T-rich codons, possibly as an adaptation to their host environments, where efficient protein production is critical. Interestingly, the TTA codon for leucine was the only codon used for this amino acid across the genus, suggesting a specialized adaptation. This preference may help the bacteria optimize their energy use and nutrient intake in the host, where rapid adaptation is necessary for survival.

The dominance of TTA and other preferred codons in Staphylococcus genomes highlights an evolutionary strategy where the bacteria have adapted their codon usage to maximize efficiency in protein synthesis. This adaptation, especially the preference for T-rich codons, seems to enhance their ability to survive and thrive in host environments, reflecting the complex interplay between microbial evolution and host-driven selective pressures**.**

Table 3a: Species-Specific Codon and Codon Pair Usage in Coagulase Positive species

Species	Most preferred Codon	RSCU	Extremely avoided Codon	Value	Most preferred codon pair	RSCPU	Extremely avoided codon pair	RSCPU
Staphylococcus aureus	TTA(Leu)	3.55	CGG(Arg)	0.08	CGG-CGG	2.991765083	TCG-ATC	-3.234867064
Staphylococcus simiae CCM 7213 = CCUG 51256	TTA(Leu)	3.87	CGG(Arg)	0.05	CGG-CCC	3.815302379	GCC-ACC	-3.596715221

Staphylococcus argenteus str. 58113	TTA(Leu)	3.64	CCC(Pro)	0.11	CGG-CTG	2.238325994	TCC-GCG	-3.247264031
Staphylococcus pseudintermedius	CGT(Arg)	3.01	CCC(Pro)	0.21	AGG-AGG	2.226601206	TCC-TGG	-3.912916074
Staphylococcus intermedius NCTC 11048	CGT(Arg)	2.92	AGG(Arg)	0.17	AAG-CGG	2.303092853	ACT-CTG	-3.201625579
Staphylococcus delphini	CGT(Arg)	3.08	AGG(Arg)	0.11	AGG-AGG	3.429406495	ACT-CTG	-3.851633566
Staphylococcus lutrae	TTA(Leu)	2.72	AGG(Arg)	0.15	AGA-AGG	1.816143949	ACT-CTC	-3.756115952

Table 3b: Species-Specific Codon and Codon Pair Usage in Coagulase -variables

Species	Most preferred Codon	RSCU	Extremely avoided Codon	Value	Most preferred codon pair	RSCPU	Extremely avoided codon pair	RSCPU
Staphylococcus hyicus	TTA(Leu)	3.2	AGG(Arg)	0.18	AGG-AGG	3.129169243	GGT-AGG	-3.551798505
Staphylococcus agnetis	TTA(Leu)	3.25	CGG(Arg)	0.19	AGG-AGG	3.100527772	CCC-TGG	-3.559172558

Table 3c: Species-Specific Codon and Codon Pair Usage in Coagulase negative species

Species	Most preferred Codon	RSCU	Extremely avoided Codon	Value	Most preferred codon pair	RSCPU	Extremely avoided codon pair	RSCPU
Staphylococcus epidermidis	TTA(Leu)	3.35	CGG(Arg)	0.09	AGG-AGG	2.216434063	CTG-GCG	-3.263784806

Staphylococcus haemolyticus	TTA(Leu)	3.49	CGG(Arg)	0.07	CGG-TGC	2.59889119	CTC-GGG	-3.71725104
Staphylococcus hominis	TTA(Leu)	3.68	CGG(Arg)	0.07	CGG-CTG	2.536495143	CTC-CGT	-3.801183764
Staphylococcus capitis	TTA(Leu)	3.4	CGG(Arg)	0.07	TCG-CTC	2.22166818	GAT-CTG	-1.995530893
Staphylococcus pettenkoferi	CGT(Arg)	2.94	AGG(Arg)	0.09	AGG-AGG	2.609979836	GGG-TCC	-3.544922628
Staphylococcus simulans	TTA(Leu)	3.43	CGG(Arg)	0.06	AGG-AGG	3.620473 18	GCG-TCC	-4.230033049
Staphylococcus warneri	TTA(Leu)	3.75	CGG(Arg)	0.05	CGG-TCG	2.669114855	GCC-AGC	-4.08292695
Staphylococcus auricularis	TTA(Leu)	2.94	AGG(Arg)	0.09	AGG-AGG	2.32581241	CTC-CTG	-3.891937698
Staphylococcus devriesei	TTA(Leu)	3.6	CTG(Leu)	0.11	CGG-TCG	2.532422529	ACC-AGA	-3.589652294
Staphylococcus pasteuri	TTA(Leu)	3.67	CGG(Arg)	0.1	CCC-CGG	3.373641732	TAC-CCC	-3.234698977
Staphylococcus chromogenes	TTA(Leu)	3.03	AGG(Arg)	0.15	AGG-AGG	3.333777307	CCC-TGG	-4.036188334
Staphylococcus saccharolyticus	TTA(Leu)	3.88	CGG(Arg)	0.04	AGG-TGC	2.55538278	GCT-CCC	-3.996505461
Staphylococcus edaphicus str. CCM 8730	TTA(Leu)	3.34	CGG(Arg)	0.12	CGG-CGG	2.160673371	AGG-ATC	-3.171982215

Staphylococcus petrasii subsp. croceilyticus	TTA(Leu)	3.71	CGG(Arg)	0.03	CGG-TCC	6.06414395	ACG-ATC	-3.697268134
Staphylococcus massiliensis S46	TTA(Leu)	3.11	CGG(Arg)	0.07	TGC-CGG	2.682471806	CTC-CCG	-3.689761336
Staphylococcus schleiferi	TTA(Leu)	3.32	CGG(Arg)	0.15	AGG-TGC	1.936091975	CCC-TGG	-4.157039377
Staphylococcus condimenti	TTA(Leu)	3.08	AGG(Arg)	0.12	AGG-AGG	3.345629379	CTT-CGG	-2.869913695
Staphylococcus piscifermentans	TTA(Leu)	3.43	CGG(Arg)	0.08	CAG-TGC	1.713939289	AGG-ATC	-3.766752269
Staphylococcus carnosus subsp. carnosus	TTA(Leu)	3.76	CGG(Arg)	0.04	CGG-CGG	4.140942958	CTC-TGG	-3.593974072
Staphylococcus arlettae	TTA(Leu)	3.64	CGG(Arg)	0.11	AGG-TGC	2.123130866	CGA-GCC	-4.098158299
Staphylococcus kloosii	TTA(Leu)	3.8	CTG(Leu)	0.12	AGG-TGC	2.755137293	AGC-CGC	-3.604448606
Staphylococcus muscae	CGT(Arg)	3.15	AGG(Arg)	0.11	AGG-AGG	3.471733861	CTA-CAG	-4.805672669
Staphylococcus rostri	CGT(Arg)	3.28	CGG(Arg)	0.1	AGG-AGG	4.135735735	ACT-CCG	-4.167851714
Staphylococcus microti	CGT(Arg)	3.2	CGG(Arg)	0.1	AAG-CGG	2.499342997	GCC-TGG	-4.050814971
Staphylococcus felis	TTA(Leu)	3.07	CGG(Arg)	0.17	AGG-CGG	2.218054564	CCT-CCG	-3.620075776

Staphylococcus caprae M23864:W1	TTA(Leu)	3.36	CGG(Arg)	0.07	AGG-AGG	3.512639492	CGA-TCG	-2.974190975
Staphylococcus saprophyticus	TTA(Leu)	3.48	CGG(Arg)	0.11	CGG-AGG	2.378650 13	ACT-CTC	-4.424719955
Staphylococcus gallinarum	TTA(Leu)	3.48	CGG(Arg)	0.08	CGG-CGG	2.311740925	AGG-ATC	-4.272664857
Staphylococcus equorum	TTA(Leu)	3.4	CGG(Arg)	0.08	AAG-CGG	1.733205147	ACT-CTC	-3.516808539
Staphylococcus xylosus	TTA(Leu)	3.46	CGG(Arg)	0.09	CCC-CGG	2.138506677	CCC-AGT	-1.965608844
Staphylococcus succinus	TTA(Leu)	3.21	CGG(Arg)	0.16	CGG-CTG	2.541774002	ACT-CTT	-2.866430 93
Staphylococcus nepalensis	TTA(Leu)	3.53	CGG(Arg)	0.13	CCG-CCC	1.921539728	ACT-CTT	-3.329884012
Staphylococcus pseudoxylosus str. S04009	TTA(Leu)	3.45	CGG(Arg)	0.09	CGG-CGG	2.013037349	AGA-TCC	-3.368486474
Staphylococcus cohnii	TTA(Leu)	3.48	CGG(Arg)	0.1	CGG-AGG	2.509058 72	ACT-CTC	-3.380199227
Staphylococcus lugdunensis	TTA(Leu)	3.34	CGG(Arg)	0.16	CGG-CTG	1.824445268	GCT-CTC	-3.430255 43
Staphylococcus sciuri	TTA(Leu)	3.68	CGG(Arg)	0.05	CCC-CCC	2.629783407	ACT-CCC	-3.555246 59
Staphylococcus lentus	TTA(Leu)	3.73	CGG(Arg)	0.05	TCC-CGG	2.271340142	TAC-CTG	-3.738080204

Staphylococcus fleurettii	TTA(Leu)	3.63	CGG(Arg)	0.05	CAG-CCC	2.097797586	ACT-CCC	-3.791425302
Staphylococcus vitulinus	TTA(Leu)	3.59	CGG(Arg)	0.05	CCC-CGG	6.442676932	CAG-GGC	-4.466729466

5.7. Exploring Codon Context Variations at Junctions

We analysed codon pair usage across the Staphylococcus genus, focusing on the patterns of dinucleotide pairs at codon pair junctions. Our analysis revealed that certain codon pairs, specifically NNG-CNN and NNA-CNN, were highly overrepresented. This suggests a strong prevalence of GpC and ApC dinucleotide pairs at these junctions, indicating that these dinucleotide combinations are favored within Staphylococcus genomes. Such overrepresentation likely plays a role in optimizing gene expression and the translation of proteins, possibly contributing to the bacteria's growth, adaptation, and survival. On the other hand, codon pairs like NNT-CNN and NNG-GNN were significantly underrepresented, pointing to a reduced presence of TpC and GpG dinucleotide pairs at codon junctions. This pattern suggests that these particular dinucleotides are less favorable, perhaps due to their association with less efficient translation or lower accuracy during protein synthesis. The underrepresentation of these dinucleotides could reflect evolutionary pressures that discourage their use, favoring combinations that contribute to more efficient protein production.

The dinucleotide patterns we observed at codon pair junctions mirror the broader trends across the genomes of the Staphylococcus genus. GpC and ApC, which were highly expressed at codon pair junctions, also emerged as the most frequently expressed dinucleotide pairs overall. This suggests that these dinucleotides play an important role in facilitating efficient translation and other essential cellular functions, helping Staphylococcus to thrive in various environments, including host organisms. In contrast, TpC and GpG were the least expressed dinucleotide pairs, both in the overall genome and at the junctions of underrepresented codon pairs. Their low frequency suggests that they may be less advantageous, potentially slowing down translation or increasing the likelihood of errors during protein synthesis. The scarcity of these dinucleotide pairs points to selective pressures that minimize their use, likely because

they do not support the same level of efficiency or accuracy as more common pairs like GpC and ApC. Our findings highlight the balance between mutational forces and natural selection in shaping the codon usage patterns of Staphylococcus. The preference for certain dinucleotides, like GpC and ApC, at key junctions, combined with the avoidance of others like TpC and GpG, suggests that the genome has evolved to favor combinations that optimize protein synthesis. This codon and dinucleotide bias play a crucial role in supporting the bacteria's ability to survive, adapt, and maintain virulence in various environments.

5.8. Exploring Synonymous Codon Similarity in *Staphylococcal* Genomes

We performed a codon adaptation index (CAI) study to evaluate the relative similarity or divergence between genes of interest and highly expressed genes within Staphylococcus genomes. With values ranging from 0 to 1, the CAI is a recognized metric that assesses how closely a gene's codon usage matches that of a reference group of highly expressed genes. Stronger codon bias is reflected in higher CAI values, which show that the genes of interest's codon use patterns resemble those of highly expressed genes more closely. This might imply that the genes are tuned for effective translation. possibly highly expressed (PHX) genes were those that were within the top 10% of CAI values in this study, whereas possibly weakly expressed (PLX) genes were those that fell within the bottom 10%. The ideal codons that are preferentially employed in highly expressed genes are highlighted by the difference in mean relative synonymous codon use (RSCU), or ΔRSCU, between PHX and PLX genes.

Our analysis revealed that CoNS (coagulase-negative staphylococci) exhibited the highest average CAI value (0.63038), suggesting a greater bias toward optimal codon usage compared to CoPS (coagulase-positive staphylococci), which had a CAI of 0.588103, and co-variable species, which had the lowest CAI value of 0.56749 (**Table 4A-AV**). These findings align with earlier observations indicating a significant codon usage bias (CUB) across the Staphylococcus genus. The high CAI values in CoNS suggest that this group has evolved a more pronounced preference for specific synonymous codons, likely to enhance translational efficiency and gene expression. In contrast, the lower CAI values in CoPS and co-variable species may indicate a more relaxed codon usage pattern, reflecting different evolutionary pressures or functional

constraints on their genomes. Additionally, the study uncovered a notable correlation between CAI values and genomic G-C content. Specifically, CoNS exhibited a higher CAI despite having a lower G-C content relative to CoPS and co-variable species. This suggests that while CoNS genomes have a reduced G-C content, they exhibit stronger selective pressures on codon usage, optimizing their synonymous codon choices to improve the efficiency of translation. This optimization could be particularly important for maintaining efficient protein synthesis under diverse environmental conditions, which may contribute to the adaptability and pathogenicity of CoNS. Further analysis revealed significant correlations between CAI values and the first two principal axes (Axis1 and Axis2) of RSCU, indicating that codon adaptation is closely tied to patterns of synonymous codon usage across the genus. The stronger bias observed in CoNS suggests that this group has developed codon usage preferences that maximize translational efficiency, potentially giving it an evolutionary advantage. In comparison, the more relaxed codon usage observed in CoPS and co-variable species may reflect a different balance between mutational pressures and natural selection. The CAI analysis provides important insights into the codon usage patterns across Staphylococcus species, highlighting the significant divergence in codon optimization between CoNS and other groups. The higher CAI values in CoNS suggest that this group has undergone more stringent selection for efficient codon usage, potentially contributing to its survival and virulence. This relationship between codon bias and genomic features, such as G-C content, underscores the role of evolutionary pressures in shaping the codon usage landscape within the Staphylococcus genus.

Table 4a: Codon usage data of *Staphylococcus agnetis*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.53*	1.07	1.65	**GCT(Ala)**	**1.01**	1.28	0.79
TTC(Phe)	0.47	0.93	0.35	GCC(Ala)	0.42	0.20	0.59
TTA(Leu)	*3.25*	4.11	2.82	**GCA(Ala)**	**1.66**	1.80	1.53
TTG(Leu)	0.77	0.30	0.99	GCG(Ala)	0.91	0.72	1.09
CTT(Leu)	0.95	1.08	0.89	TAT(Tyr)	*1.54*	1.21	1.58
CTC(Leu)	0.40	0.19	0.53	*TAC(Tyr)*	0.46	0.79	0.42
CTA(Leu)	0.44	0.27	0.47	CAT(His)	*1.52*	1.17	1.56
CTG(Leu)	0.20	0.05	0.31	**CAC(His)**	0.48	0.83	0.44
ATT(Ile)	*1.82*	1.93	1.58	*CAA(Gln)*	*1.78*	1.93	1.64
ATC(Ile)	0.64	0.98	0.66	CAG(Gln)	0.22	0.07	0.36
ATA(Ile)	0.54	0.09	0.76	AAT(Asn)	*1.45*	1.07	1.49
GTT(Val)	*1.30*	1.66	1.04	*AAC(Asn)*	0.55	0.93	0.51
GTC(Val)	0.65	0.34	0.83	*AAA(Lys)*	*1.69*	1.81	1.56
GTA(Val)	*1.19*	1.48	1.05	AAG(Lys)	0.31	0.19	0.44
GTG(Val)	0.87	0.51	1.08	GAT(Asp)	*1.50*	1.33	1.47
TCT(Ser)	*1.46*	1.90	1.17	**GAC(Asp)**	0.50	0.67	0.53
TCC(Ser)	0.32	0.10	0.49	*GAA(Glu)*	*1.65*	1.76	1.53
TCA(Ser)	*1.82*	2.36	1.58	GAG(Glu)	0.35	0.24	0.47
TCG(Ser)	0.45	0.18	0.70	*TGT(Cys)*	*1.65*	1.74	1.55
AGT(Ser)	*1.49*	0.98	1.54	TGC(Cys)	0.35	0.26	0.45
AGC(Ser)	0.46	0.47	0.51	**CGT(Arg)**	**2.94**	4.39	2.09
CCT(Pro)	**1.48**	1.45	1.33	CGC(Arg)	0.94	0.68	1.02
CCC(Pro)	0.24	0.03	0.44	CGA(Arg)	0.79	0.17	1.09
CCA(Pro)	**1.69**	2.26	1.38	CGG(Arg)	0.19	0.00	0.48
CCG(Pro)	0.59	0.26	0.84	AGA(Arg)	0.99	0.75	0.98
ACT(Thr)	0.85	1.21	0.64	AGG(Arg)	0.16	0.00	0.34
ACC(Thr)	0.31	0.06	0.53	**GGT(Gly)**	**1.92**	2.71	1.42
ACA(Thr)	*1.84*	2.11	1.64	GGC(Gly)	0.71	0.52	0.85
ACG(Thr)	**1.01**	0.62	1.20	GGA(Gly)	0.88	0.60	0.89
				GGG(Gly)	0.50	0.18	0.84

Table 4b: Codon usage data of *Staphylococcus argenteus*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	1.49	1.10	1.61	GCT(Ala)	1.27	1.55	1.13
TTC(Phe)	0.51	0.90	0.39	GCC(Ala)	0.27	0.11	0.37
TTA(Leu)	3.64	4.51	3.02	GCA(Ala)	1.91	1.97	1.74
TTG(Leu)	0.84	0.37	1.23	GCG(Ala)	0.55	0.36	0.75
CTT(Leu)	0.72	0.69	0.69	TAT(Tyr)	1.59	1.35	1.61
CTC(Leu)	0.12	0.02	0.22	TAC(Tyr)	0.41	0.65	0.39
CTA(Leu)	0.54	0.38	0.61	CAT(His)	1.62	1.33	1.64
CTG(Leu)	0.13	0.02	0.23	CAC(His)	0.38	0.67	0.36
ATT(Ile)	1.86	2.00	1.63	CAA(Gln)	1.77	1.94	1.62
ATC(Ile)	0.48	0.83	0.41	CAG(Gln)	0.23	0.06	0.38
ATA(Ile)	0.66	0.17	0.97	AAT(Asn)	1.54	1.28	1.58
GTT(Val)	1.66	1.92	1.41	AAC(Asn)	0.46	0.72	0.42
GTC(Val)	0.4	0.16	0.54	AAA(Lys)	1.66	1.83	1.49
GTA(Val)	1.41	1.64	1.27	AAG(Lys)	0.34	0.17	0.51
GTG(Val)	0.53	0.28	0.77	GAT(Asp)	1.59	1.46	1.60
TCT(Ser)	1.26	1.45	1.12	GAC(Asp)	0.41	0.54	0.40
TCC(Ser)	0.15	0.03	0.23	GAA(Glu)	1.69	1.81	1.52
TCA(Ser)	2.03	2.64	1.71	GAG(Glu)	0.31	0.19	0.48
TCG(Ser)	0.38	0.09	0.67	TGT(Cys)	1.6	1.66	1.50
AGT(Ser)	1.69	1.25	1.80	TGC(Cys)	0.4	0.34	0.50
AGC(Ser)	0.49	0.54	0.48	CGT(Arg)	2.3	3.75	1.55
CCT(Pro)	1.34	1.41	1.16	CGC(Arg)	0.51	0.45	0.56
CCC(Pro)	0.11	0.01	0.19	CGA(Arg)	0.85	0.22	1.27
CCA(Pro)	2.06	2.38	1.90	CGG(Arg)	0.07	0.00	0.19
CCG(Pro)	0.49	0.20	0.75	AGA(Arg)	2.04	1.56	1.88
ACT(Thr)	1.18	1.60	1.01	AGG(Arg)	0.23	0.01	0.55
ACC(Thr)	0.16	0.03	0.27	GGT(Gly)	2.15	2.68	1.73
ACA(Thr)	2.04	2.09	1.80	GGC(Gly)	0.58	0.48	0.66
ACG(Thr)	0.62	0.28	0.93	GGA(Gly)	0.98	0.74	1.08
				GGG(Gly)	0.29	0.10	0.53

Table 4c: Codon usage data of *Staphylococcus arlettae*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.53*	1.14	1.66	**GCT(Ala)**	**1.37**	1.65	1.18
TTC(Phe)	0.47	0.86	0.34	GCC(Ala)	0.34	0.15	0.46
TTA(Leu)	*3.64*	4.42	3.19	**GCA(Ala)**	**1.79**	1.89	1.68
TTG(Leu)	0.81	0.38	1.13	GCG(Ala)	0.50	0.31	0.68
CTT(Leu)	0.59	0.60	0.53	TAT(Tyr)	*1.58*	1.41	1.59
CTC(Leu)	0.20	0.04	0.28	*TAC(Tyr)*	0.42	0.59	0.41
CTA(Leu)	0.59	0.52	0.57	CAT(His)	*1.54*	1.33	1.58
CTG(Leu)	0.17	0.05	0.30	**CAC(His)**	0.46	0.67	0.42
ATT(Ile)	*1.78*	1.92	1.66	*CAA(Gln)*	*1.76*	1.93	1.58
ATC(Ile)	0.49	0.81	0.46	CAG(Gln)	0.24	0.07	0.42
ATA(Ile)	0.73	0.27	0.88	AAT(Asn)	*1.50*	1.26	1.59
GTT(Val)	*1.35*	1.66	1.20	*AAC(Asn)*	0.50	0.74	0.41
GTC(Val)	0.48	0.26	0.58	*AAA(Lys)*	*1.69*	1.86	1.50
GTA(Val)	*1.53*	1.77	1.34	AAG(Lys)	0.31	0.14	0.50
GTG(Val)	0.63	0.31	0.88	GAT(Asp)	*1.55*	1.43	1.54
TCT(Ser)	*1.34*	1.59	1.19	**GAC(Asp)**	0.45	0.57	0.46
TCC(Ser)	0.25	0.07	0.34	*GAA(Glu)*	*1.69*	1.82	1.53
TCA(Ser)	*1.83*	2.60	1.59	GAG(Glu)	0.31	0.18	0.47
TCG(Ser)	0.34	0.06	0.60	*TGT(Cys)*	*1.67*	1.78	1.59
AGT(Ser)	*1.75*	1.25	1.78	TGC(Cys)	0.33	0.22	0.41
AGC(Ser)	0.48	0.43	0.50	**CGT(Arg)**	**2.56**	4.00	1.96
CCT(Pro)	**1.40**	1.35	1.32	CGC(Arg)	0.65	0.41	0.74
CCC(Pro)	0.18	0.03	0.33	**CGA(Arg)**	0.75	0.15	1.15
CCA(Pro)	**1.97**	2.46	1.68	CGG(Arg)	0.11	0.00	0.30
CCG(Pro)	0.45	0.16	0.67	AGA(Arg)	*1.73*	1.43	1.47
ACT(Thr)	0.58	0.28	1.13	AGG(Arg)	0.19	0.01	0.39
ACC(Thr)	0.23	0.05	0.36	**GGT(Gly)**	**2.13**	2.59	1.82
ACA(Thr)	*1.86*	2.06	1.72	GGC(Gly)	0.68	0.62	0.72
ACG(Thr)	0.58	0.28	0.78	GGA(Gly)	0.89	0.71	0.95
				GGG(Gly)	0.31	0.08	0.51

Table 4d: Codon usage data of *Staphylococcus aureus*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.74***	1.20	1.54	**GCT(Ala)**	**1.26**	1.57	1.03
TTC(Phe)	0.53	0.80	0.46	GCC(Ala)	0.28	0.13	0.44
TTA(Leu)	***3.55***	4.43	2.65	**GCA(Ala)**	**1.87**	1.91	1.78
TTG(Leu)	0.89	0.44	1.22	GCG(Ala)	0.59	0.39	0.75
CTT(Leu)	0.70	0.67	0.66	TAT(Tyr)	***1.57***	1.40	1.53
CTC(Leu)	0.14	0.02	0.28	***TAC(Tyr)***	0.43	0.60	0.47
CTA(Leu)	0.56	0.41	0.65	CAT(His)	***1.60***	1.40	1.51
CTG(Leu)	0.17	0.04	0.53	**CAC(His)**	0.40	0.60	0.49
ATT(Ile)	***1.83***	2.00	1.61	***CAA(Gln)***	***1.76***	1.91	1.59
ATC(Ile)	0.51	0.79	0.46	CAG(Gln)	0.24	0.09	0.41
ATA(Ile)	0.66	0.21	0.93	AAT(Asn)	***1.53***	1.40	1.45
GTT(Val)	***1.63***	1.93	1.34	***AAC(Asn)***	0.47	0.60	0.55
GTC(Val)	0.43	0.18	0.65	***AAA(Lys)***	***1.63***	1.81	1.42
GTA(Val)	***1.37***	1.60	1.20	AAG(Lys)	0.37	0.19	0.58
GTG(Val)	0.57	0.29	0.80	GAT(Asp)	***1.57***	1.46	1.51
TCT(Ser)	***1.29***	0.29	0.75	**GAC(Asp)**	0.43	0.54	0.49
TCC(Ser)	0.17	0.02	0.44	***GAA(Glu)***	***1.68***	1.80	1.50
TCA(Ser)	***1.98***	2.53	1.58	GAG(Glu)	0.32	0.20	0.50
TCG(Ser)	0.38	0.11	0.64	***TGT(Cys)***	***1.61***	1.69	1.50
AGT(Ser)	***1.68***	1.31	1.85	TGC(Cys)	0.39	0.31	0.50
AGC(Ser)	0.51	0.55	0.61	**CGT(Arg)**	**2.25**	3.54	1.33
CCT(Pro)	**1.36**	1.47	0.69	CGC(Arg)	0.56	0.47	0.90
CCC(Pro)	0.12	0.01	0.27	CGA(Arg)	0.83	0.28	1.20
CCA(Pro)	**2.02**	2.27	1.92	CGG(Arg)	0.08	0.00	0.25
CCG(Pro)	0.50	0.24	0.69	AGA(Arg)	***2.02***	1.68	1.75
ACT(Thr)	***1.15***	0.32	0.88	AGG(Arg)	0.25	0.03	0.57
ACC(Thr)	0.19	0.03	0.34	**GGT(Gly)**	**2.14**	2.65	1.63
ACA(Thr)	***2.00***	2.13	1.84	GGC(Gly)	0.62	0.51	0.74
ACG(Thr)	0.66	0.32	0.88	GGA(Gly)	0.94	0.75	1.06
				GGG(Gly)	0.29	0.09	0.57

Table 4e: Codon usage data of *Staphylococcus auricularis*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.22***	0.72	1.44	**GCT(Ala)**	**1.02**	1.42	0.81
TTC(Phe)	0.78	1.28	0.56	GCC(Ala)	0.39	0.23	0.47
TTA(Leu)	***2.94***	3.80	2.45	**GCA(Ala)**	**1.67**	1.93	1.49
TTG(Leu)	0.75	0.31	1.08	GCG(Ala)	0.92	0.42	1.22
CTT(Leu)	0.78	1.19	0.61	TAT(Tyr)	***1.40***	1.03	1.52
CTC(Leu)	0.79	0.32	0.92	***TAC(Tyr)***	0.60	0.97	0.48
CTA(Leu)	0.47	0.36	0.47	CAT(His)	***1.33***	0.90	1.48
CTG(Leu)	0.27	0.03	0.46	CAC(His)	0.67	1.10	0.52
ATT(Ile)	***1.52***	1.32	1.50	***CAA(Gln)***	***1.72***	1.97	1.46
ATC(Ile)	***1.11***	1.60	0.95	CAG(Gln)	0.28	0.03	0.54
ATA(Ile)	0.38	0.08	0.54	AAT(Asn)	***1.27***	0.85	1.43
GTT(Val)	***1.02***	1.59	0.80	***AAC(Asn)***	0.73	1.15	0.57
GTC(Val)	0.93	0.48	1.10	***AAA(Lys)***	***1.73***	1.96	1.40
GTA(Val)	***1.11***	1.49	0.85	AAG(Lys)	0.27	0.04	0.60
GTG(Val)	0.93	0.44	1.25	GAT(Asp)	***1.46***	1.25	1.54
TCT(Ser)	***1.45***	1.80	1.17	**GAC(Asp)**	0.54	0.75	0.46
TCC(Ser)	0.31	0.06	0.48	***GAA(Glu)***	***1.71***	1.93	1.51
TCA(Ser)	***1.84***	2.40	1.52	GAG(Glu)	0.29	0.07	0.49
TCG(Ser)	0.32	0.01	0.75	TGT(Cys)	***1.55***	1.39	1.52
AGT(Ser)	***1.43***	0.94	1.48	**TGC(Cys)**	0.45	0.61	0.48
AGC(Ser)	0.64	0.79	0.60	**CGT(Arg)**	**2.69**	3.86	1.94
CCT(Pro)	**1.31**	1.35	1.21	CGC(Arg)	**1.02**	0.82	0.93
CCC(Pro)	0.24	0.05	0.40	CGA(Arg)	0.68	0.09	1.16
CCA(Pro)	**1.74**	2.42	1.33	CGG(Arg)	0.22	0.00	0.57
CCG(Pro)	0.71	0.17	1.05	***AGA(Arg)***	***1.31***	1.22	1.18
ACT(Thr)	0.95	1.50	0.61	AGG(Arg)	0.09	0.00	0.23
ACC(Thr)	0.42	0.09	0.61	**GGT(Gly)**	**2.07**	2.77	1.54
ACA(Thr)	***1.79***	2.23	1.34	GGC(Gly)	0.80	0.67	0.85
ACG(Thr)	0.95	0.18	1.43	GGA(Gly)	0.73	0.48	0.84
				GGG(Gly)	0.39	0.08	0.76

Table 4f: Codon usage data of *Staphylococcus capiti*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.38***	1.13	1.42	**GCT(Ala)**	**1.58**	1.87	1.22
TTC(Phe)	0.62	0.87	0.58	GCC(Ala)	0.31	0.15	0.45
TTA(Leu)	***3.40***	4.34	2.67	**GCA(Ala)**	**1.62**	1.68	1.64
TTG(Leu)	0.59	0.22	0.88	GCG(Ala)	0.50	0.31	0.69
CTT(Leu)	0.98	0.85	1.00	***TAT(Tyr)***	***1.55***	1.55	1.49
CTC(Leu)	0.27	0.09	0.46	TAC(Tyr)	0.45	0.45	0.51
CTA(Leu)	0.64	0.48	0.72	***CAT(His)***	***1.54***	1.45	1.46
CTG(Leu)	0.12	0.02	0.26	**CAC(His)**	0.46	0.55	0.54
ATT(Ile)	***1.75***	2.02	1.54	***CAA(Gln)***	***1.76***	1.90	1.59
ATC(Ile)	0.56	0.69	0.63	CAG(Gln)	0.24	0.10	0.41
ATA(Ile)	0.69	0.29	0.83	***AAT(Asn)***	***1.52***	1.42	1.42
GTT(Val)	***1.59***	1.85	1.29	***AAC(Asn)***	0.48	0.58	0.58
GTC(Val)	0.44	0.21	0.65	***AAA(Lys)***	***1.63***	1.82	1.45
GTA(Val)	***1.40***	1.63	1.25	AAG(Lys)	0.37	0.18	0.55
GTG(Val)	0.56	0.30	0.81	***GAT(Asp)***	***1.57***	1.58	1.47
TCT(Ser)	***1.55***	1.84	1.30	GAC(Asp)	0.43	0.42	0.53
TCC(Ser)	0.21	0.05	0.36	***GAA(Glu)***	***1.63***	1.80	1.48
TCA(Ser)	***1.87***	2.21	1.54	GAG(Glu)	0.37	0.20	0.52
TCG(Ser)	0.25	0.07	0.54	***TGT(Cys)***	***1.60***	1.60	1.48
AGT(Ser)	***1.66***	1.44	1.66	TGC(Cys)	0.40	0.40	0.52
AGC(Ser)	0.46	0.40	0.61	**CGT(Arg)**	**2.30**	3.24	1.99
CCT(Pro)	**1.77**	1.78	1.54	CGC(Arg)	0.53	0.49	0.69
CCC(Pro)	0.19	0.04	0.40	CGA(Arg)	0.73	0.34	0.98
CCA(Pro)	**1.73**	2.02	1.56	CGG(Arg)	0.07	0.01	0.20
CCG(Pro)	0.31	0.16	0.50	***AGA(Arg)***	***2.18***	1.91	1.71
ACT(Thr)	***1.52***	1.96	1.13	AGG(Arg)	0.20	0.03	0.42
ACC(Thr)	0.24	0.06	0.45	**GGT(Gly)**	**2.07**	2.61	1.73
ACA(Thr)	***1.71***	1.72	1.65	GGC(Gly)	0.57	0.46	0.71
ACG(Thr)	0.53	0.26	0.77	GGA(Gly)	**1.06**	0.83	1.15
				GGG(Gly)	0.29	0.10	0.42

Table 4g: Codon usage data of *Staphylococcus caprae*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.31*	0.89	1.46	**GCT(Ala)**	**1.40**	1.81	1.17
TTC(Phe)	0.69	1.11	0.54	GCC(Ala)	0.35	0.15	0.49
TTA(Leu)	*3.36*	4.42	2.78	**GCA(Ala)**	**1.65**	1.72	1.52
TTG(Leu)	0.64	0.23	1.02	GCG(Ala)	0.60	0.32	0.81
CTT(Leu)	0.95	0.82	0.86	TAT(Tyr)	*1.53*	1.31	1.58
CTC(Leu)	0.33	0.08	0.48	*TAC(Tyr)*	0.47	0.69	0.42
CTA(Leu)	0.57	0.43	0.60	CAT(His)	*1.53*	1.18	1.59
CTG(Leu)	0.15	0.03	0.27	**CAC(His)**	0.47	0.82	0.41
ATT(Ile)	*1.78*	1.80	1.61	*CAA(Gln)*	*1.78*	1.95	1.62
ATC(Ile)	0.63	1.07	0.57	CAG(Gln)	0.22	0.05	0.38
ATA(Ile)	0.59	0.13	0.83	AAT(Asn)	*1.49*	1.17	1.55
GTT(Val)	*1.46*	1.80	1.17	*AAC(Asn)*	0.51	0.83	0.45
GTC(Val)	0.55	0.23	0.76	*AAA(Lys)*	*1.62*	1.84	1.40
GTA(Val)	*1.32*	1.65	1.14	AAG(Lys)	0.38	0.16	0.60
GTG(Val)	0.67	0.32	0.93	GAT(Asp)	*1.52*	1.39	1.54
TCT(Ser)	*1.51*	1.86	1.28	**GAC(Asp)**	0.48	0.61	0.46
TCC(Ser)	0.24	0.05	0.37	*GAA(Glu)*	*1.67*	1.84	1.49
TCA(Ser)	*1.84*	2.37	1.61	GAG(Glu)	0.33	0.16	0.51
TCG(Ser)	0.28	0.05	0.59	TGT(Cys)	*1.58*	1.53	1.56
AGT(Ser)	*1.65*	1.17	1.64	**TGC(Cys)**	0.42	0.47	0.44
AGC(Ser)	0.49	0.50	0.51	**CGT(Arg)**	**2.33**	3.81	1.73
CCT(Pro)	**1.71**	1.69	1.54	CGC(Arg)	0.54	0.50	0.56
CCC(Pro)	0.16	0.03	0.35	CGA(Arg)	0.73	0.14	1.19
CCA(Pro)	**1.75**	2.16	1.51	CGG(Arg)	0.07	0.00	0.24
CCG(Pro)	0.37	0.12	0.60	AGA(Arg)	*2.14*	1.53	1.88
ACT(Thr)	*1.29*	1.93	0.95	AGG(Arg)	0.19	0.02	0.40
ACC(Thr)	0.25	0.03	0.45	**GGT(Gly)**	**2.00**	2.66	1.58
ACA(Thr)	*1.78*	1.78	1.62	GGC(Gly)	0.67	0.57	0.75
ACG(Thr)	0.67	0.26	0.97	GGA(Gly)	**1.03**	0.70	1.15
				GGG(Gly)	0.30	0.07	0.52

Table 4h: Codon usage data of *Staphylococcus carnosus*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.31*	1.16	1.32	**GCT(Ala)**	**1.50**	2.07	1.01
TTC(Phe)	0.69	0.84	0.68	GCC(Ala)	0.34	0.13	0.44
TTA(Leu)	*3.76*	4.64	3.18	**GCA(Ala)**	**1.52**	1.70	1.27
TTG(Leu)	0.56	0.24	0.87	GCG(Ala)	0.64	0.10	1.28
CTT(Leu)	0.86	0.71	0.82	TAT(Tyr)	*1.55*	1.48	1.55
CTC(Leu)	0.18	0.05	0.32	*TAC(Tyr)*	0.45	0.52	0.45
CTA(Leu)	0.53	0.34	0.61	CAT(His)	*1.53*	1.34	1.55
CTG(Leu)	0.11	0.02	0.20	**CAC(His)**	0.47	0.66	0.45
ATT(Ile)	*1.89*	1.98	1.82	*CAA(Gln)*	*1.80*	1.94	1.68
ATC(Ile)	0.62	0.81	0.64	CAG(Gln)	0.20	0.06	0.32
ATA(Ile)	0.49	0.21	0.54	AAT(Asn)	*1.49*	1.39	1.45
GTT(Val)	*1.40*	1.69	1.07	*AAC(Asn)*	0.51	0.61	0.55
GTC(Val)	0.44	0.22	0.66	*AAA(Lys)*	*1.69*	1.88	1.48
GTA(Val)	*1.53*	1.79	1.27	*AAG(Lys)*	0.31	1.39	0.52
GTG(Val)	0.63	0.31	1.01	*GAT(Asp)*	*1.54*	1.49	1.49
TCT(Ser)	*1.50*	1.66	1.32	**GAC(Asp)**	0.46	0.51	0.51
TCC(Ser)	0.15	0.04	0.22	*GAA(Glu)*	*1.69*	1.83	1.51
TCA(Ser)	*1.92*	2.35	1.63	GAG(Glu)	0.31	0.17	0.49
TCG(Ser)	0.26	0.06	0.52	*TGT(Cys)*	*1.60*	1.60	1.59
AGT(Ser)	*1.64*	1.37	1.73	TGC(Cys)	0.40	0.40	0.41
AGC(Ser)	0.53	0.52	0.58	**CGT(Arg)**	**2.76**	3.38	2.62
CCT(Pro)	**1.67**	1.73	1.46	CGC(Arg)	0.70	0.54	0.94
CCC(Pro)	0.14	0.04	0.23	CGA(Arg)	0.58	0.26	0.79
CCA(Pro)	**1.85**	2.12	1.68	CGG(Arg)	0.04	0.00	0.12
CCG(Pro)	0.34	0.11	0.63	*AGA(Arg)*	*1.79*	1.79	1.29
ACT(Thr)	*1.37*	1.83	0.94	AGG(Arg)	0.13	0.02	0.23
ACC(Thr)	0.23	0.04	0.40	**GGT(Gly)**	**2.07**	2.59	1.74
ACA(Thr)	*1.76*	1.87	1.61	GGC(Gly)	0.78	0.65	0.99
ACG(Thr)	0.64	0.25	1.05	GGA(Gly)	0.86	0.65	0.85
				GGG(Gly)	0.29	0.11	0.42

Table 4i: Codon usage data of *Staphylococcus chromogenes*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.50***	1.50	1.55	GCT(Ala)	**1.00**	0.97	1.02
TTC(Phe)	0.50	0.50	0.45	GCC(Ala)	0.59	0.57	0.67
TTA(Leu)	***3.03***	2.95	2.99	**GCA(Ala)**	**1.49**	1.49	1.45
TTG(Leu)	0.83	0.85	0.82	**GCG(Ala)**	0.92	0.98	0.87
CTT(Leu)	0.96	1.00	0.99	***TAT(Tyr)***	***1.50***	1.52	1.52
CTC(Leu)	0.53	0.54	0.56	***TAC(Tyr)***	0.50	0.48	0.48
CTA(Leu)	0.44	0.44	0.42	CAT(His)	***1.49***	1.47	1.54
CTG(Leu)	0.21	0.22	0.22	**CAC(His)**	0.51	0.53	0.46
ATT(Ile)	***1.88***	1.87	1.90	***CAA(Gln)***	***1.78***	1.77	1.75
ATC(Ile)	0.66	0.65	0.66	CAG(Gln)	0.22	0.23	0.25
ATA(Ile)	0.46	0.47	0.44	AAT(Asn)	***1.40***	1.40	1.42
GTT(Val)	***1.24***	1.21	1.22	***AAC(Asn)***	0.60	0.60	0.58
GTC(Val)	0.83	0.83	0.84	AAA(Lys)	***1.69***	1.67	1.68
GTA(Val)	0.99	0.92	0.99	***AAG(Lys)***	0.31	0.33	0.32
GTG(Val)	0.95	1.04	0.95	GAT(Asp)	***1.46***	1.44	1.50
TCT(Ser)	***1.46***	1.37	1.50	**GAC(Asp)**	0.54	0.56	0.50
TCC(Ser)	0.39	0.42	0.41	GAA(Glu)	***1.63***	1.61	1.62
TCA(Ser)	***1.72***	1.72	1.71	**GAG(Glu)**	0.37	0.39	0.38
TCG(Ser)	0.43	0.45	0.39	***TGT(Cys)***	***1.61***	1.65	1.56
AGT(Ser)	***1.49***	1.52	1.48	TGC(Cys)	0.39	0.35	0.44
AGC(Ser)	0.51	0.53	0.53	**CGT(Arg)**	**2.80**	2.78	2.73
CCT(Pro)	**1.54**	1.45	1.62	**CGC(Arg)**	0.90	0.90	0.85
CCC(Pro)	0.32	0.31	0.34	**CGA(Arg)**	0.83	0.85	0.83
CCA(Pro)	**1.51**	1.59	1.46	CGG(Arg)	0.23	0.20	0.24
CCG(Pro)	0.63	0.65	0.58	AGA(Arg)	***1.08***	1.10	1.22
ACT(Thr)	0.87	0.81	0.94	**AGG(Arg)**	0.15	0.17	0.13
ACC(Thr)	0.36	0.37	0.37	GGT(Gly)	**1.57**	1.52	1.53
ACA(Thr)	***1.83***	1.86	1.76	GGC(Gly)	0.78	0.79	0.82
ACG(Thr)	0.94	0.96	0.93	GGA(Gly)	**1.03**	1.02	1.03
				GGG(Gly)	0.61	0.67	0.62

Table 4j: Codon usage data of *Staphylococcus cohnii*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.54*	1.18	1.57	**GCT(Ala)**	**1.35**	1.62	1.20
TTC(Phe)	0.46	0.82	0.43	GCC(Ala)	0.28	0.11	0.43
TTA(Leu)	*3.48*	4.21	2.84	**GCA(Ala)**	**1.87**	1.93	1.77
TTG(Leu)	0.83	0.41	1.09	GCG(Ala)	0.50	0.34	0.61
CTT(Leu)	0.73	0.79	0.76	*TAT(Tyr)*	*1.62*	1.56	1.48
CTC(Leu)	0.21	0.05	0.37	TAC(Tyr)	0.38	0.44	0.52
CTA(Leu)	0.57	0.48	0.64	CAT(His)	*1.57*	1.33	1.52
CTG(Leu)	0.18	0.05	0.30	**CAC(His)**	0.43	0.67	0.48
ATT(Ile)	*1.69*	1.84	1.48	*CAA(Gln)*	*1.75*	1.93	1.47
ATC(Ile)	0.53	0.84	0.54	CAG(Gln)	0.25	0.07	0.53
ATA(Ile)	0.78	0.32	0.98	AAT(Asn)	*1.53*	1.30	1.40
GTT(Val)	*1.52*	1.76	1.35	*AAC(Asn)*	0.47	0.70	0.60
GTC(Val)	0.44	0.24	0.55	*AAA(Lys)*	*1.70*	1.87	1.48
GTA(Val)	*1.44*	1.68	1.24	AAG(Lys)	0.30	0.13	0.52
GTG(Val)	0.61	0.32	0.86	*GAT(Asp)*	*1.59*	1.52	1.33
TCT(Ser)	*1.43*	1.62	1.18	GAC(Asp)	0.41	0.48	0.67
TCC(Ser)	0.24	0.07	0.34	*GAA(Glu)*	*1.71*	1.84	1.37
TCA(Ser)	*1.90*	2.56	1.62	GAG(Glu)	0.29	0.16	0.63
TCG(Ser)	0.30	0.06	0.49	*TGT(Cys)*	*1.61*	1.73	1.46
AGT(Ser)	*1.65*	1.23	1.64	TGC(Cys)	0.39	0.27	0.54
AGC(Ser)	0.48	0.46	0.73	**CGT(Arg)**	**2.20**	3.79	1.56
CCT(Pro)	**1.47**	1.35	1.51	CGC(Arg)	0.56	0.46	0.65
CCC(Pro)	0.19	0.04	0.36	CGA(Arg)	0.63	0.17	1.00
CCA(Pro)	**1.95**	2.40	1.60	CGG(Arg)	0.10	0.00	0.23
CCG(Pro)	0.39	0.21	0.53	AGA(Arg)	*2.26*	1.57	2.00
ACT(Thr)	*1.19*	1.51	1.01	AGG(Arg)	0.26	0.01	0.56
ACC(Thr)	0.25	0.06	0.33	**GGT(Gly)**	**2.14**	2.56	1.66
ACA(Thr)	*1.97*	2.12	1.88	GGC(Gly)	0.66	0.57	0.94
ACG(Thr)	0.59	0.31	0.77	GGA(Gly)	0.91	0.79	0.94
				GGG(Gly)	0.28	0.08	0.47

Table 4k: Codon usage data of *Staphylococcus condimenti*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.23*	1.02	1.28	**GCT(Ala)**	**1.29**	1.74	1.28
TTC(Phe)	0.77	0.98	0.72	GCC(Ala)	0.30	0.17	0.31
TTA(Leu)	*3.08*	3.52	2.98	**GCA(Ala)**	**1.89**	1.95	1.81
TTG(Leu)	*1.11*	1.00	1.19	GCG(Ala)	0.52	0.42	0.59
CTT(Leu)	0.85	0.88	0.84	TAT(Tyr)	*1.56*	1.46	1.57
CTC(Leu)	0.25	0.15	0.28	*TAC(Tyr)*	0.44	0.54	0.43
CTA(Leu)	0.37	0.29	0.37	*CAT(His)*	*1.54*	1.43	1.52
CTG(Leu)	0.33	0.17	0.34	**CAC(His)**	0.46	0.57	0.48
ATT(Ile)	*1.77*	1.76	1.76	*CAA(Gln)*	*1.73*	1.84	1.67
ATC(Ile)	0.76	0.98	0.73	CAG(Gln)	0.27	0.16	0.33
ATA(Ile)	0.47	0.27	0.51	AAT(Asn)	*1.42*	1.24	1.43
GTT(Val)	*1.43*	1.60	1.33	*AAC(Asn)*	0.58	0.76	0.57
GTC(Val)	0.52	0.35	0.56	*AAA(Lys)*	*1.70*	1.82	1.67
GTA(Val)	*1.45*	1.62	1.44	AAG(Lys)	0.30	0.18	0.33
GTG(Val)	0.60	0.42	0.67	GAT(Asp)	*1.56*	1.48	1.54
TCT(Ser)	*1.54*	1.79	1.63	**GAC(Asp)**	0.44	0.52	0.46
TCC(Ser)	0.21	0.11	0.26	*GAA(Glu)*	*1.73*	1.81	1.72
TCA(Ser)	*1.84*	2.04	1.68	GAG(Glu)	0.27	0.19	0.28
TCG(Ser)	0.30	0.16	0.36	*TGT(Cys)*	*1.35*	1.30	1.26
AGT(Ser)	*1.40*	1.17	1.43	TGC(Cys)	0.65	0.70	0.74
AGC(Ser)	0.70	0.72	0.64	**CGT(Arg)**	**2.65**	3.38	2.81
CCT(Pro)	**1.52**	1.57	1.50	CGC(Arg)	0.89	0.87	0.90
CCC(Pro)	0.15	0.07	0.16	CGA(Arg)	0.64	0.36	0.65
CCA(Pro)	**1.50**	1.76	1.39	CGG(Arg)	0.13	0.05	0.12
CCG(Pro)	0.82	0.60	0.94	AGA(Arg)	*1.56*	1.32	1.37
ACT(Thr)	*1.26*	1.51	1.20	AGG(Arg)	0.12	0.03	0.15
ACC(Thr)	0.23	0.09	0.29	**GGT(Gly)**	**1.83**	2.10	1.84
ACA(Thr)	*2.06*	2.11	2.00	**GGC(Gly)**	0.83	0.85	0.78
ACG(Thr)	0.45	0.29	0.51	GGA(Gly)	**1.06**	0.88	1.07
				GGG(Gly)	0.28	0.16	0.31

Table 4l: Codon usage data of *Staphylococcus delphini*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.39*	0.89	1.52	**GCT(Ala)**	0.77	1.03	0.60
TTC(Phe)	0.61	1.11	0.48	GCC(Ala)	0.41	0.18	0.52
TTA(Leu)	*2.91*	4.14	2.23	**GCA(Ala)**	**1.65**	1.94	1.46
TTG(Leu)	*1.11*	0.50	1.52	GCG(Ala)	**1.17**	0.85	1.42
CTT(Leu)	0.71	0.93	0.57	TAT(Tyr)	*1.43*	1.05	1.56
CTC(Leu)	0.64	0.21	0.87	*TAC(Tyr)*	0.57	0.95	0.44
CTA(Leu)	0.30	0.14	0.36	CAT(His)	*1.46*	1.02	1.52
CTG(Leu)	0.34	0.07	0.45	**CAC(His)**	0.54	0.98	0.48
ATT(Ile)	*1.98*	1.82	1.86	*CAA(Gln)*	*1.73*	1.94	1.54
ATC(Ile)	0.73	1.15	0.67	CAG(Gln)	0.27	0.06	0.46
ATA(Ile)	0.28	0.03	0.47	AAT(Asn)	*1.27*	0.82	1.42
GTT(Val)	0.98	1.42	0.72	AAC(Asn)	0.73	1.18	0.58
GTC(Val)	0.96	0.47	1.13	*AAA(Lys)*	*1.63*	1.84	1.44
GTA(Val)	0.91	1.37	0.77	AAG(Lys)	0.37	0.16	0.56
GTG(Val)	**1.14**	0.75	1.38	GAT(Asp)	*1.36*	1.16	1.40
TCT(Ser)	*1.10*	1.51	0.91	**GAC(Asp)**	0.64	0.84	0.60
TCC(Ser)	0.28	0.05	0.47	*GAA(Glu)*	*1.64*	1.76	1.50
TCA(Ser)	*1.87*	2.72	1.33	GAG(Glu)	0.36	0.24	0.50
TCG(Ser)	0.64	0.16	1.05	*TGT(Cys)*	*1.65*	1.76	1.54
AGT(Ser)	*1.54*	0.70	1.72	TGC(Cys)	0.35	0.24	0.80
AGC(Ser)	0.56	0.70	0.52	**CGT(Arg)**	**3.08**	4.49	2.19
CCT(Pro)	**1.15**	1.19	1.01	CGC(Arg)	0.97	0.67	1.12
CCC(Pro)	0.21	0.03	0.37	CGA(Arg)	0.70	0.11	1.15
CCA(Pro)	**1.64**	2.42	1.21	CGG(Arg)	0.23	0.00	0.66
CCG(Pro)	1.00	0.36	1.42	*AGA(Arg)*	0.91	0.72	0.67
ACT(Thr)	0.71	1.10	0.53	AGG(Arg)	0.11	0.00	0.22
ACC(Thr)	0.26	0.04	0.47	**GGT(Gly)**	**1.93**	2.73	1.43
ACA(Thr)	*1.87*	2.37	1.44	GGC(Gly)	0.90	0.66	1.01
ACG(Thr)	**1.16**	0.49	1.57	GGA(Gly)	0.63	0.44	0.72
				GGG(Gly)	0.54	0.17	0.84

Table 4m: Codon usage data of *Staphylococcus devriesei*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon (Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.37*	0.93	1.47	**GCT(Ala)**	**1.50**	1.90	1.16
TTC(Phe)	0.63	1.07	0.53	GCC(Ala)	0.41	0.16	0.60
TTA(Leu)	*3.60*	4.64	3.03	**GCA(Ala)**	**1.48**	1.61	1.40
TTG(Leu)	0.57	0.17	0.81	GCG(Ala)	0.61	0.33	0.84
CTT(Leu)	0.85	0.75	0.79	TAT(Tyr)	*1.53*	1.29	1.52
CTC(Leu)	0.21	0.03	0.35	*TAC(Tyr)*	0.47	0.71	0.48
CTA(Leu)	0.68	0.40	0.82	CAT(His)	*1.52*	1.18	1.51
CTG(Leu)	0.11	0.01	0.20	**CAC(His)**	0.48	0.82	0.49
ATT(Ile)	*1.83*	1.81	1.67	*CAA(Gln)*	*1.80*	1.97	1.63
ATC(Ile)	0.60	1.10	0.55	CAG(Gln)	0.20	0.03	0.37
ATA(Ile)	0.56	0.08	0.79	AAT(Asn)	*1.46*	1.11	1.48
GTT(Val)	*1.41*	1.78	1.08	*AAC(Asn)*	0.54	0.89	0.52
GTC(Val)	0.46	0.21	0.65	*AAA(Lys)*	*1.66*	1.89	1.44
GTA(Val)	*1.54*	1.77	1.38	AAG(Lys)	0.34	0.11	0.56
GTG(Val)	0.59	0.24	0.89	GAT(Asp)	*1.51*	1.30	1.51
TCT(Ser)	*1.45*	1.73	1.19	**GAC(Asp)**	0.49	0.70	0.49
TCC(Ser)	0.18	0.02	0.32	*GAA(Glu)*	*1.68*	1.88	1.46
TCA(Ser)	*1.86*	2.57	1.63	GAG(Glu)	0.32	0.12	0.54
TCG(Ser)	0.30	0.04	0.58	*TGT(Cys)*	*1.60*	1.65	1.53
AGT(Ser)	*1.63*	1.09	1.59	TGC(Cys)	0.40	0.35	0.47
AGC(Ser)	0.58	0.54	0.70	**CGT(Arg)**	**2.70**	4.06	1.96
CCT(Pro)	**1.63**	1.56	1.42	CGC(Arg)	0.64	0.50	0.80
CCC(Pro)	0.16	0.01	0.33	CGA(Arg)	0.63	0.08	1.05
CCA(Pro)	**1.84**	2.34	1.57	CGG(Arg)	0.09	0.00	0.24
CCG(Pro)	0.37	0.09	0.68	AGA(Arg)	*1.81*	1.35	1.65
ACT(Thr)	*1.44*	2.03	1.01	AGG(Arg)	0.14	0.00	0.30
ACC(Thr)	0.26	0.03	0.49	**GGT(Gly)**	**1.98**	2.61	1.61
ACA(Thr)	*1.68*	1.77	1.50	GGC(Gly)	0.80	0.75	0.89
ACG(Thr)	0.62	0.17	1.00	GGA(Gly)	0.89	0.57	0.99
				GGG(Gly)	0.32	0.07	0.52

Table 4n: Codon usage data of *Staphylococcus edaphicus*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.49*	1.17	1.57	**GCT(Ala)**	**1.25**	1.60	1.02
TTC(Phe)	0.51	0.83	0.43	GCC(Ala)	0.35	0.17	0.45
TTA(Leu)	*3.34*	4.20	2.87	**GCA(Ala)**	**1.85**	1.87	1.77
TTG(Leu)	0..86	0.41	1.13	GCG(Ala)	0.55	0.36	0.76
CTT(Leu)	0.75	0.78	0.72	TAT(Tyr)	*1.60*	1.47	1.63
CTC(Leu)	0.22	0.06	0.31	*TAC(Tyr)*	0.40	0.53	0.37
CTA(Leu)	0.62	0.49	0.66	CAT(His)	*1.56*	1.29	1.59
CTG(Leu)	0.20	0.05	0.30	**CAC(His)**	0.44	0.71	0.41
ATT(Ile)	*1.68*	1.83	1.54	*CAA(Gln)*	*1.72*	1.91	1.55
ATC(Ile)	0.54	0.83	0.53	CAG(Gln)	0.28	0.09	0.45
ATA(Ile)	0.78	0.34	0.93	AAT(Asn)	*1.54*	1.30	1.57
GTT(Val)	*1.46*	1.81	1.24	*AAC(Asn)*	0.46	0.70	0.43
GTC(Val)	0.47	0.25	0.59	*AAA(Lys)*	*1.65*	1.83	1.45
GTA(Val)	*1.43*	1.60	1.29	AAG(Lys)	0.35	0.17	0.55
GTG(Val)	0.64	0.33	0.88	GAT(Asp)	*1.59*	1.50	1.57
TCT(Ser)	*1.34*	1.56	1.14	**GAC(Asp)**	0.41	0.50	0.43
TCC(Ser)	0.27	0.10	0.37	*GAA(Glu)*	*1.68*	1.80	1.56
TCA(Ser)	*1.87*	2.47	1.64	GAG(Glu)	0.32	0.20	0.44
TCG(Ser)	0.35	0.13	0.56	*TGT(Cys)*	*1.62*	1.76	1.48
AGT(Ser)	*1.64*	1.29	1.71	TGC(Cys)	0.38	0.24	0.52
AGC(Ser)	0.53	0.46	0.57	**CGT(Arg)**	**2.18**	3.67	1.54
CCT(Pro)	**1.36**	1.37	1.25	CGC(Arg)	0.61	0.49	0.65
CCC(Pro)	0.24	0.04	0.38	CGA(Arg)	0.62	0.15	0.93
CCA(Pro)	**1.96**	2.33	1.69	CGG(Arg)	0.12	0.01	0.28
CCG(Pro)	0.44	0.25	0.67	AGA(Arg)	*2.19*	1.65	2.04
ACT(Thr)	*1.10*	1.53	0.86	AGG(Arg)	0.28	0.03	0.56
ACC(Thr)	0.29	0.07	0.44	**GGT(Gly)**	**2.06**	2.53	1.79
ACA(Thr)	*1.92*	2.09	1.66	GGC(Gly)	0.71	0.63	0.77
ACG(Thr)	0.69	0.31	1.03	GGA(Gly)	0.90	0.74	0.92
				GGG(Gly)	0.33	0.11	0.52

Table 4o: Codon usage data of *Staphylococcus epidermidis*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.51*	1.12	1.60	**GCT(Ala)**	**1.53**	1.76	1.24
TTC(Phe)	0.49	0.88	0.40	GCC(Ala)	0.30	0.14	0.44
TTA(Leu)	*3.35*	4.21	2.98	**GCA(Ala)**	**1.73**	1.80	1.66
TTG(Leu)	0.69	0.31	0.96	GCG(Ala)	0.44	0.30	0.66
CTT(Leu)	0.95	0.88	0.85	TAT(Tyr)	*1.59*	1.42	1.63
CTC(Leu)	0.25	0.07	0.34	*TAC(Tyr)*	0.41	0.58	0.37
CTA(Leu)	0.62	0.49	0.62	CAT(His)	*1.58*	1.34	1.58
CTG(Leu)	0.13	0.03	0.24	**CAC(His)**	0.42	0.66	0.42
ATT(Ile)	*1.72*	1.91	1.53	*CAA(Gln)*	*1.76*	1.92	1.56
ATC(Ile)	0.49	0.83	0.46	CAG(Gln)	**1.76**	0.08	0.44
ATA(Ile)	0.79	0.26	1.01	AAT(Asn)	*1.57*	1.34	1.59
GTT(Val)	*1.63*	1.87	1.43	*AAC(Asn)*	0.43	0.66	0.41
GTC(Val)	0.44	0.21	0.59	*AAA(Lys)*	*1.65*	1.84	1.47
GTA(Val)	*1.37*	1.59	1.16	AAG(Lys)	0.35	0.16	0.53
GTG(Val)	0.56	0.33	0.83	GAT(Asp)	*1.61*	1.50	1.59
TCT(Ser)	*1.51*	1.77	1.29	**GAC(Asp)**	0.39	0.50	0.41
TCC(Ser)	0.24	0.10	0.37	*GAA(Glu)*	*1.65*	1.82	1.49
TCA(Ser)	*1.84*	2.42	1.55	GAG(Glu)	0.35	0.18	0.51
TCG(Ser)	0.30	0.07	0.54	*TGT(Cys)*	*1.64*	1.65	1.62
AGT(Ser)	*1.68*	1.24	1.77	TGC(Cys)	0.36	0.35	0.38
AGC(Ser)	0.43	0.40	0.49	**CGT(Arg)**	**2.19**	3.53	1.84
CCT(Pro)	**1.69**	1.59	1.57	**CGC(Arg)**	0.45	0.50	0.46
CCC(Pro)	0.21	0.03	0.40	**CGA(Arg)**	0.81	0.24	1.11
CCA(Pro)	**1.75**	2.23	1.58	CGG(Arg)	0.09	0.00	0.18
CCG(Pro)	0.34	0.15	0.45	AGA(Arg)	*2.23*	1.72	1.90
ACT(Thr)	*1.40*	1.77	1.15	AGG(Arg)	0.23	0.01	0.51
ACC(Thr)	0.25	0.05	0.46	**GGT(Gly)**	**2.03**	2.53	1.66
ACA(Thr)	*1.85*	1.95	1.66	GGC(Gly)	0.49	0.45	0.50
ACG(Thr)	0.49	0.22	0.73	GGA(Gly)	**1.16**	0.91	1.31
				GGG(Gly)	0.31	0.11	0.53

Table 4p: Codon usage data of *Staphylococcus equorum*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.45***	1.11	1.51	**GCT(Ala)**	**1.30**	1.58	1.10
TTC(Phe)	0.55	0.89	0.49	GCC(Ala)	0.29	0.13	0.41
TTA(Leu)	***3.40***	4.13	2.80	**GCA(Ala)**	**1.87**	1.95	1.73
TTG(Leu)	0.77	0.38	1.06	GCG(Ala)	0.54	0.34	0.77
CTT(Leu)	0.81	0.85	0.80	TAT(Tyr)	***1.56***	1.42	1.53
CTC(Leu)	0.25	0.06	0.39	***TAC(Tyr)***	0.44	0.58	0.47
CTA(Leu)	0.59	0.53	0.59	CAT(His)	***1.52***	1.30	1.52
CTG(Leu)	0.18	0.04	0.37	**CAC(His)**	0.48	0.70	0.48
ATT(Ile)	***1.66***	1.84	1.44	***CAA(Gln)***	***1.74***	1.92	1.52
ATC(Ile)	0.54	0.87	0.50	CAG(Gln)	0.26	0.08	0.48
ATA(Ile)	0.80	0.29	1.06	AAT(Asn)	***1.48***	1.24	1.51
GTT(Val)	***1.45***	1.80	1.22	***AAC(Asn)***	0.52	0.76	0.49
GTC(Val)	0.43	0.21	0.57	***AAA(Lys)***	***1.67***	1.86	1.46
GTA(Val)	0.62	1.67	1.35	AAG(Lys)	0.33	0.14	0.54
GTG(Val)	0.62	0.31	0.86	GAT(Asp)	***1.56***	1.45	1.51
TCT(Ser)	***1.41***	1.57	1.22	**GAC(Asp)**	0.44	0.55	0.49
TCC(Ser)	0.22	0.06	0.35	***GAA(Glu)***	***1.69***	1.82	1.51
TCA(Ser)	***1.84***	2.53	1.52	GAG(Glu)	0.31	0.18	0.49
TCG(Ser)	0.33	0.13	0.55	***TGT(Cys)***	***1.64***	1.73	1.53
AGT(Ser)	***1.70***	1.27	1.75	TGC(Cys)	0.36	0.27	0.47
AGC(Ser)	0.52	0.44	0.62	**CGT(Arg)**	**2.26**	3.67	1.57
CCT(Pro)	**1.44**	1.33	1.34	CGC(Arg)	0.58	0.53	0.62
CCC(Pro)	0.17	0.04	0.34	CGA(Arg)	0.60	0.17	0.96
CCA(Pro)	**1.96**	2.37	1.70	CGG(Arg)	0.08	0.00	0.25
CCG(Pro)	0.43	0.26	0.63	AGA(Arg)	***2.24***	1.62	2.03
ACT(Thr)	***1.18***	1.52	0.96	AGG(Arg)	0.24	0.00	0.56
ACC(Thr)	0.22	0.04	0.37	**GGT(Gly)**	**2.04**	2.43	1.62
ACA(Thr)	***1.97***	2.10	1.83	GGC(Gly)	0.75	0.75	0.86
ACG(Thr)	0.63	0.34	0.84	GGA(Gly)	0.89	0.71	1.00
				GGG(Gly)	0.31	0.11	0.52

Table 4q: Codon usage data of *Staphylococcus felis*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.57***	1.16	1.65	**GCT(Ala)**	**1.23**	1.50	1.03
TTC(Phe)	0.43	0.84	0.35	GCC(Ala)	0.35	0.17	0.45
TTA(Leu)	***3.07***	3.95	2.57	**GCA(Ala)**	**1.76**	1.88	1.63
TTG(Leu)	0.88	0.44	1.26	GCG(Ala)	0.66	0.45	0.89
CTT(Leu)	0.94	1.06	0.86	TAT(Tyr)	***1.55***	1.33	1.59
CTC(Leu)	0.35	0.12	0.46	***TAC(Tyr)***	0.45	0.67	0.41
CTA(Leu)	0.53	0.36	0.52	CAT(His)	***1.53***	1.22	1.57
CTG(Leu)	0.22	0.06	0.33	**CAC(His)**	0.47	0.78	0.43
ATT(Ile)	***1.77***	1.88	1.56	***CAA(Gln)***	***1.72***	1.91	1.55
ATC(Ile)	0.58	0.94	0.57	CAG(Gln)	0.28	0.09	0.45
ATA(Ile)	0.65	0.18	0.87	AAT(Asn)	***1.50***	1.18	1.57
GTT(Val)	***1.48***	1.77	1.14	***AAC(Asn)***	0.50	0.82	0.43
GTC(Val)	0.64	0.35	0.80	***AAA(Lys)***	***1.59***	1.77	1.42
GTA(Val)	***1.18***	1.44	1.04	AAG(Lys)	0.41	0.23	0.58
GTG(Val)	0.71	0.44	1.01	GAT(Asp)	***1.54***	1.39	1.56
TCT(Ser)	***1.33***	1.57	1.13	**GAC(Asp)**	0.46	0.61	0.44
TCC(Ser)	0.26	0.07	0.43	***GAA(Glu)***	***1.57***	1.73	1.38
TCA(Ser)	***1.94***	2.61	1.68	GAG(Glu)	0.43	0.27	0.62
TCG(Ser)	0.45	0.17	0.72	***TGT(Cys)***	***1.58***	1.72	1.49
AGT(Ser)	***1.52***	1.05	1.59	TGC(Cys)	0.42	0.28	0.51
AGC(Ser)	0.50	0.54	0.45	**CGT(Arg)**	**2.45**	4.07	1.80
CCT(Pro)	**1.53**	1.40	1.39	CGC(Arg)	0.75	0.70	0.82
CCC(Pro)	**1.70**	0.05	0.41	CGA(Arg)	**1.01**	0.31	1.34
CCA(Pro)	**1.70**	2.28	1.43	CGG(Arg)	0.17	0.00	0.40
CCG(Pro)	0.54	0.27	0.78	AGA(Arg)	***1.38***	0.91	1.21
ACT(Thr)	***1.02***	1.37	0.79	AGG(Arg)	0.24	0.01	0.44
ACC(Thr)	0.25	0.05	0.38	**GGT(Gly)**	**1.78**	2.50	1.38
ACA(Thr)	***2.01***	2.18	1.81	GGC(Gly)	0.64	0.49	0.70
ACG(Thr)	0.71	0.40	1.01	GGA(Gly)	**1.09**	0.80	1.15
				GGG(Gly)	0.49	0.20	0.76

Table 4r: Codon usage data of *Staphylococcus fleurettii*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.43*	1.07	1.55	*GCT(Ala)*	**1.50**	1.73	1.26
TTC(Phe)	0.57	0.93	0.45	GCC(Ala)	0.19	0.08	0.35
TTA(Leu)	*3.63*	4.36	2.94	*GCA(Ala)*	**1.94**	1.96	1.83
TTG(Leu)	0.67	0.30	0.97	GCG(Ala)	0.37	0.22	0.57
CTT(Leu)	0.81	0.83	0.83	TAT(Tyr)	*1.60*	1.42	1.56
CTC(Leu)	0.17	0.02	0.35	*TAC(Tyr)*	0.40	0.58	0.44
CTA(Leu)	0.59	0.47	0.67	CAT(His)	*1.60*	1.41	1.61
CTG(Leu)	0.12	0.02	0.23	*CAC(His)*	0.40	0.59	0.39
ATT(Ile)	*1.61*	1.80	1.48	*CAA(Gln)*	*1.79*	1.95	1.57
ATC(Ile)	0.45	0.81	0.41	CAG(Gln)	0.21	0.05	0.43
ATA(Ile)	0.94	0.39	1.10	AAT(Asn)	*1.52*	1.21	1.57
GTT(Val)	*1.63*	1.86	1.38	*AAC(Asn)*	0.48	0.79	0.43
GTC(Val)	0.34	0.15	0.51	*AAA(Lys)*	*1.68*	1.87	1.47
GTA(Val)	*1.61*	1.79	1.40	AAG(Lys)	0.32	0.13	0.53
GTG(Val)	0.42	0.20	0.71	GAT(Asp)	*1.59*	1.48	1.54
TCT(Ser)	*1.51*	1.66	1.42	*GAC(Asp)*	0.41	0.52	0.46
TCC(Ser)	0.17	0.03	0.29	*GAA(Glu)*	*1.74*	1.85	1.57
TCA(Ser)	*2.03*	2.76	1.50	GAG(Glu)	0.26	0.15	0.43
TCG(Ser)	0.24	0.04	0.50	*TGT(Cys)*	*1.67*	1.84	1.49
AGT(Ser)	*1.62*	1.14	1.76	TGC(Cys)	0.33	0.16	0.51
AGC(Ser)	0.43	0.38	0.53	*CGT(Arg)*	**1.63**	3.40	1.29
CCT(Pro)	**1.49**	1.44	1.50	CGC(Arg)	0.30	0.31	0.46
CCC(Pro)	0.10	0.01	0.23	CGA(Arg)	0.56	0.15	0.91
CCA(Pro)	**2.09**	2.44	1.71	CGG(Arg)	0.05	0.00	0.15
CCG(Pro)	0.33	0.10	0.56	AGA(Arg)	*3.20*	2.13	2.68
ACT(Thr)	*1.35*	1.68	1.09	AGG(Arg)	0.26	0.01	0.51
ACC(Thr)	0.13	0.01	0.32	*GGT(Gly)*	**2.07**	2.55	1.61
ACA(Thr)	*2.02*	2.08	1.80	GGC(Gly)	0.50	0.45	0.63
ACG(Thr)	0.50	0.22	0.79	GGA(Gly)	**1.17**	0.93	1.31
				GGG(Gly)	0.26	0.08	0.45

Table 4s: Codon usage data of *Staphylococcus gallinarum*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.48***	1.06	1.61	**GCT(Ala)**	**1.24**	1.58	1.06
TTC(Phe)	0.52	0.94	0.39	GCC(Ala)	0.33	0.14	0.43
TTA(Leu)	***3.48***	4.33	2.94	**GCA(Ala)**	**1.92**	1.95	1.79
TTG(Leu)	0.90	0.46	1.26	GCG(Ala)	0.51	0.32	0.72
CTT(Leu)	0.65	0.72	0.59	TAT(Tyr)	***1.61***	1.41	1.63
CTC(Leu)	0.21	0.04	0.28	***TAC(Tyr)***	0.39	0.59	0.37
CTA(Leu)	0.59	0.42	0.63	CAT(His)	***1.55***	1.26	1.63
CTG(Leu)	0.17	0.03	0.29	**CAC(His)**	0.45	0.74	0.37
ATT(Ile)	***1.69***	1.78	1.54	***CAA(Gln)***	***1.73***	1.94	1.54
ATC(Ile)	0.57	1.04	0.45	CAG(Gln)	0.27	0.06	0.46
ATA(Ile)	0.74	0.17	1.01	AAT(Asn)	***1.54***	1.22	1.64
GTT(Val)	***1.44***	1.81	1.22	***AAC(Asn)***	0.46	0.78	0.36
GTC(Val)	0.47	0.22	0.58	***AAA(Lys)***	***1.52***	1.77	1.29
GTA(Val)	***1.49***	1.67	0.58	AAG(Lys)	0.48	0.23	0.71
GTG(Val)	0.60	0.29	0.84	GAT(Asp)	***1.59***	1.43	1.62
TCT(Ser)	***1.33***	1.55	1.16	**GAC(Asp)**	0.41	0.57	0.38
TCC(Ser)	0.21	0.03	0.33	***GAA(Glu)***	***1.66***	1.79	1.46
TCA(Ser)	***1.92***	2.66	1.62	GAG(Glu)	0.34	0.21	0.54
TCG(Ser)	0.35	0.08	0.65	***TGT(Cys)***	***1.64***	1.75	1.57
AGT(Ser)	***1.73***	1.22	1.81	TGC(Cys)	0.36	0.25	0.43
AGC(Ser)	0.46	0.46	0.44	**CGT(Arg)**	**2.17**	3.63	1.66
CCT(Pro)	**1.35**	1.30	1.26	CGC(Arg)	0.56	0.53	0.56
CCC(Pro)	0.16	0.03	0.27	**CGA(Arg)**	0.74	0.24	1.10
CCA(Pro)	**2.05**	2.50	1.76	CGG(Arg)	0.08	0	0.23
CCG(Pro)	0.44	0.17	0.70	AGA(Arg)	***2.20***	1.58	1.89
ACT(Thr)	***1.18***	1.60	1.01	AGG(Arg)	0.26	0.01	0.57
ACC(Thr)	0.26	0.04	0.39	**GGT(Gly)**	**2.13**	2.67	1.56
ACA(Thr)	***1.94***	2.08	1.70	GGC(Gly)	0.60	0.50	0.61
ACG(Thr)	0.61	0.28	0.90	GGA(Gly)	0.97	0.73	1.03
				GGG(Gly)	0.30	0.09	0.51

Table 4t: Codon usage data of *Staphylococcus haemolyticus*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.34***	0.92	1.50	**GCT(Ala)**	**1.29**	1.72	1.02
TTC(Phe)	0.66	1.08	0.50	GCC(Ala)	0.36	0.15	0.56
TTA(Leu)	***3.49***	4.62	2.87	**GCA(Ala)**	**1.72**	1.77	1.53
TTG(Leu)	0.68	0.24	1.01	GCG(Ala)	0.63	0.36	0.89
CTT(Leu)	0.81	0.73	0.70	TAT(Tyr)	***1.54***	1.32	1.56
CTC(Leu)	0.24	0.05	0.41	***TAC(Tyr)***	0.46	0.68	0.44
CTA(Leu)	0.63	0.35	0.72	CAT(His)	***1.53***	1.22	1.59
CTG(Leu)	0.15	0.02	0.28	**CAC(His)**	0.47	0.78	0.41
ATT(Ile)	***1.80***	1.86	1.63	***CAA(Gln)***	***1.78***	1.96	1.59
ATC(Ile)	0.63	1.04	0.54	CAG(Gln)	0.22	0.04	0.41
ATA(Ile)	0.58	0.10	0.83	AAT(Asn)	***1.49***	1.12	1.50
GTT(Val)	***1.54***	1.88	1.24	***AAC(Asn)***	0.51	0.88	0.50
GTC(Val)	0.49	0.21	0.68	***AAA(Lys)***	***1.64***	1.89	1.38
GTA(Val)	***1.37***	1.62	1.18	AAG(Lys)	0.36	0.11	0.62
GTG(Val)	0.60	0.28	0.90	GAT(Asp)	***1.55***	1.39	1.55
TCT(Ser)	***1.38***	1.66	1.18	**GAC(Asp)**	0.45	0.61	0.45
TCC(Ser)	0.18	0.02	0.36	***GAA(Glu)***	***1.66***	1.85	1.43
TCA(Ser)	***1.98***	2.61	1.62	GAG(Glu)	0.34	0.15	0.57
TCG(Ser)	0.30	0.04	0.58	***TGT(Cys)***	***1.60***	1.65	1.48
AGT(Ser)	***1.67***	1.23	1.70	TGC(Cys)	0.40	0.35	0.52
AGC(Ser)	0.48	0.43	0.56	**CGT(Arg)**	**2.56**	4.01	1.86
CCT(Pro)	**1.51**	1.44	1.37	CGC(Arg)	0.59	0.45	0.63
CCC(Pro)	0.15	0.02	0.34	CGA(Arg)	0.82	0.15	1.35
CCA(Pro)	**1.93**	2.44	1.56	CGG(Arg)	0.07	0.00	0.25
CCG(Pro)	0.41	0.11	0.73	AGA(Arg)	***1.82***	1.39	1.55
ACT(Thr)	***1.16***	1.72	0.82	AGG(Arg)	0.15	0.00	0.36
ACC(Thr)	0.25	0.03	0.47	**GGT(Gly)**	**2.15**	2.75	1.64
ACA(Thr)	***1.84***	1.96	1.57	GGC(Gly)	0.69	0.55	0.81
ACG(Thr)	0.75	0.29	1.14	GGA(Gly)	0.87	0.63	1.00
				GGG(Gly)	0.29	0.07	0.56

Table 4u: Codon usage data of *Staphylococcus hominis*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.53***	1.11	1.61	**GCT(Ala)**	**1.53**	1.83	1.23
TTC(Phe)	0.47	0.89	0.39	GCC(Ala)	0.28	0.12	0.45
TTA(Leu)	***3.68***	4.63	3.01	**GCA(Ala)**	**1.76**	1.77	1.69
TTG(Leu)	0.55	0.20	0.88	GCG(Ala)	0.44	0.28	0.63
CTT(Leu)	0.93	0.79	0.90	TAT(Tyr)	***1.64***	1.51	1.62
CTC(Leu)	0.19	0.04	0.36	***TAC(Tyr)***	0.36	0.49	0.38
CTA(Leu)	0.54	0.32	0.63	CAT(His)	***1.65***	1.45	1.59
CTG(Leu)	0.11	0.02	0.22	**CAC(His)**	0.35	0.55	0.41
ATT(Ile)	***1.93***	2.08	1.62	***CAA(Gln)***	***1.80***	1.96	1.61
ATC(Ile)	0.45	0.76	0.44	CAG(Gln)	0.20	0.04	0.39
ATA(Ile)	0.63	0.16	0.95	AAT(Asn)	***1.61***	1.38	1.57
GTT(Val)	***1.48***	1.79	1.26	***AAC(Asn)***	0.39	0.62	0.43
GTC(Val)	0.40	0.18	0.61	***AAA(Lys)***	***1.72***	1.90	1.50
GTA(Val)	***1.55***	1.72	1.32	AAG(Lys)	0.28	0.10	0.50
GTG(Val)	0.56	0.31	0.81	GAT(Asp)	***1.64***	1.54	1.58
TCT(Ser)	***1.62***	1.84	1.41	**GAC(Asp)**	0.36	0.46	0.42
TCC(Ser)	0.16	0.02	0.27	***GAA(Glu)***	***1.71***	1.87	1.51
TCA(Ser)	***1.86***	2.37	1.55	GAG(Glu)	0.29	0.13	0.49
TCG(Ser)	0.25	0.08	0.45	***TGT(Cys)***	***1.68***	1.69	1.50
AGT(Ser)	***1.73***	1.30	1.82	TGC(Cys)	0.32	0.31	0.50
AGC(Ser)	0.38	0.39	0.50	**CGT(Arg)**	**2.51**	3.88	1.82
CCT(Pro)	**1.69**	1.57	1.67	CGC(Arg)	0.42	0.46	0.48
CCC(Pro)	0.18	0.04	0.34	CGA(Arg)	0.76	0.19	1.30
CCA(Pro)	**1.86**	2.28	1.57	CGG(Arg)	0.07	0.00	0.17
CCG(Pro)	0.27	0.11	0.41	AGA(Arg)	***2.10***	1.46	1.82
ACT(Thr)	***1.38***	1.86	1.11	AGG(Arg)	0.14	0.01	0.41
ACC(Thr)	0.23	0.04	0.37	**GGT(Gly)**	**1.99**	2.71	1.52
ACA(Thr)	***1.91***	1.91	1.79	GGC(Gly)	0.52	0.47	0.60
ACG(Thr)	0.49	0.19	0.72	GGA(Gly)	**1.22**	0.74	1.38
				GGG(Gly)	0.27	0.09	0.50

Table 4v: Codon usage data of *Staphylococcus hyicus*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.54***	1.08	1.68	**GCT(Ala)**	0.99	1.28	0.81
TTC(Phe)	0.46	0.92	0.32	GCC(Ala)	0.42	0.19	0.55
TTA(Leu)	***3.20***	4.14	2.72	**GCA(Ala)**	**1.69**	1.83	1.57
TTG(Leu)	0.82	0.32	1.12	GCG(Ala)	0.90	0.71	1.08
CTT(Leu)	0.96	1.07	0.89	TAT(Tyr)	***1.54***	1.23	1.60
CTC(Leu)	0.38	0.16	0.48	***TAC(Tyr)***	0.46	0.77	0.40
CTA(Leu)	0.44	0.26	0.51	CAT(His)	***1.51***	1.17	1.57
CTG(Leu)	0.19	0.05	0.28	**CAC(His)**	0.49	0.83	0.43
ATT(Ile)	***1.82***	1.93	1.62	***CAA(Gln)***	***1.77***	1.93	1.61
ATC(Ile)	0.61	0.94	0.56	CAG(Gln)	0.23	0.07	0.39
ATA(Ile)	0.57	0.13	0.83	AAT(Asn)	***1.45***	1.10	1.55
GTT(Val)	***1.30***	1.64	1.05	***AAC(Asn)***	0.55	0.90	0.45
GTC(Val)	0.62	0.34	0.80	***AAA(Lys)***	***1.69***	1.80	1.56
GTA(Val)	***1.20***	1.48	1.04	AAG(Lys)	0.31	0.20	0.44
GTG(Val)	0.87	0.54	1.11	GAT(Asp)	***1.50***	1.36	1.50
TCT(Ser)	***1.41***	1.86	1.14	**GAC(Asp)**	0.50	0.64	0.50
TCC(Ser)	0.32	0.10	0.40	***GAA(Glu)***	***1.65***	1.77	1.49
TCA(Ser)	***1.83***	2.33	1.64	GAG(Glu)	0.35	0.23	0.51
TCG(Ser)	0.47	0.19	0.71	***TGT(Cys)***	***1.61***	1.71	1.56
AGT(Ser)	***1.49***	1.00	1.62	TGC(Cys)	0.39	0.29	0.44
AGC(Ser)	0.47	0.52	0.49	**CGT(Arg)**	**2.83**	4.24	2.10
CCT(Pro)	**1.44**	1.45	1.27	CGC(Arg)	0.92	0.68	0.99
CCC(Pro)	0.26	0.04	0.44	CGA(Arg)	0.85	0.26	1.15
CCA(Pro)	**1.69**	2.22	1.39	CGG(Arg)	0.22	0.01	0.50
CCG(Pro)	0.62	0.29	0.90	AGA(Arg)	0.99	0.80	0.89
ACT(Thr)	0.83	1.17	0.68	AGG(Arg)	0.18	0.01	0.38
ACC(Thr)	0.31	0.08	0.45	**GGT(Gly)**	**1.85**	2.57	1.46
ACA(Thr)	***1.86***	2.13	1.64	GGC(Gly)	0.70	0.60	0.73
ACG(Thr)	0.99	0.62	1.23	GGA(Gly)	0.92	0.64	1.01
				GGG(Gly)	0.53	0.19	0.81

Table 4w: Codon usage data of *Staphylococcus intermedius*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.42*	0.92	1.55	**GCT(Ala)**	0.86	1.00	0.80
TTC(Phe)	0.58	1.08	0.45	GCC(Ala)	0.39	0.16	0.50
TTA(Leu)	*2.90*	4.00	2.29	**GCA(Ala)**	**1.71**	1.99	1.49
TTG(Leu)	*1.11*	0.55	1.43	GCG(Ala)	**1.04**	0.85	1.21
CTT(Leu)	0.76	0.95	0.68	TAT(Tyr)	*1.46*	1.12	1.54
CTC(Leu)	0.56	0.20	0.75	*TAC(Tyr)*	0.54	0.88	0.46
CTA(Leu)	0.36	0.2	0.43	CAT(His)	*1.48*	1.11	1.50
CTG(Leu)	0.30	0.08	0.42	**CAC(His)**	0.52	0.89	0.50
ATT(Ile)	*1.91*	1.86	1.75	*CAA(Gln)*	*1.73*	1.94	1.56
ATC(Ile)	0.70	1.09	0.62	CAG(Gln)	0.27	0.06	0.44
ATA(Ile)	0.39	0.04	0.63	AAT(Asn)	*1.32*	0.89	1.44
GTT(Val)	*1.09*	1.40	0.84	*AAC(Asn)*	0.68	1.11	0.56
GTC(Val)	0.87	0.46	1.03	*AAA(Lys)*	*1.62*	1.80	1.46
GTA(Val)	0.98	1.40	0.82	AAG(Lys)	0.38	0.20	0.54
GTG(Val)	**1.07**	0.74	1.31	GAT(Asp)	*1.41*	1.22	1.45
TCT(Ser)	*1.13*	1.22	0.96	**GAC(Asp)**	0.59	0.78	0.55
TCC(Ser)	0.27	0.06	0.45	*GAA(Glu)*	*1.64*	1.77	1.49
TCA(Ser)	*1.88*	2.76	1.38	GAG(Glu)	0.36	0.23	0.51
TCG(Ser)	0.58	0.27	0.91	*TGT(Cys)*	*1.61*	1.74	1.56
AGT(Ser)	*1.58*	1.03	1.73	TGC(Cys)	0.39	0.26	0.44
AGC(Ser)	0.56	0.67	0.58	**CGT(Arg)**	**2.92**	4.33	2.07
CCT(Pro)	**1.18**	1.18	1.11	CGC(Arg)	0.87	0.68	0.94
CCC(Pro)	0.23	0.04	0.39	CGA(Arg)	0.74	0.14	1.18
CCA(Pro)	**1.72**	2.41	1.25	CGG(Arg)	0.20	0.01	0.56
CCG(Pro)	0.88	0.37	1.24	AGA(Arg)	*1.10*	0.84	0.91
ACT(Thr)	0.78	1.10	0.63	AGG(Arg)	0.17	0.01	0.35
ACC(Thr)	0.26	0.04	0.45	**GGT(Gly)**	**1.96**	2.72	1.47
ACA(Thr)	*1.94*	2.36	1.57	GGC(Gly)	0.80	0.62	0.93
ACG(Thr)	*1.02*	0.50	1.36	GGA(Gly)	0.73	0.47	0.83
				GGG(Gly)	0.51	0.19	0.78

Table 4x: Codon usage data of *Staphylococcus kloosii*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.45***	1.01	1.58	**GCT(Ala)**	**1.40**	1.76	1.17
TTC(Phe)	0.55	0.99	0.42	GCC(Ala)	0.38	0.14	0.52
TTA(Leu)	***3.80***	4.47	3.37	**GCA(Ala)**	**1.68**	1.75	1.62
TTG(Leu)	0.66	0.25	0.98	GCG(Ala)	0.54	0.35	0.68
CTT(Leu)	0.66	0.76	0.57	TAT(Tyr)	***1.55***	1.32	1.60
CTC(Leu)	0.17	0.02	0.26	***TAC(Tyr)***	0.45	0.68	0.40
CTA(Leu)	0.60	0.48	0.60	CAT(His)	***1.53***	1.19	1.63
CTG(Leu)	0.12	0.02	0.21	***CAC(His)***	0.47	0.81	0.37
ATT(Ile)	***1.69***	1.78	1.56	***CAA(Gln)***	***1.81***	1.95	1.68
ATC(Ile)	0.53	1.00	0.43	CAG(Gln)	0.19	0.05	0.32
ATA(Ile)	0.77	0.23	1.01	AAT(Asn)	***1.46***	1.06	1.57
GTT(Val)	***1.51***	1.83	1.25	***AAC(Asn)***	0.54	0.94	0.43
GTC(Val)	0.46	0.18	0.59	***AAA(Lys)***	***1.70***	1.89	1.53
GTA(Val)	***1.52***	1.77	1.40	AAG(Lys)	0.30	0.11	0.47
GTG(Val)	0.51	0.21	0.76	GAT(Asp)	***1.48***	1.29	1.51
TCT(Ser)	***1.41***	1.68	1.22	GAC(Asp)	0.52	0.71	0.49
TCC(Ser)	0.19	0.04	0.28	***GAA(Glu)***	***1.73***	1.86	1.59
TCA(Ser)	***1.81***	2.59	1.41	GAG(Glu)	0.27	0.14	0.41
TCG(Ser)	0.34	0.05	0.64	***TGT(Cys)***	***1.64***	1.77	1.55
AGT(Ser)	***1.72***	1.14	1.83	TGC(Cys)	0.36	0.23	0.45
AGC(Ser)	0.54	0.50	0.62	**CGT(Arg)**	**2.40**	3.78	1.72
CCT(Pro)	**1.47**	1.42	1.31	CGC(Arg)	0.57	0.41	0.66
CCC(Pro)	0.16	0.02	0.30	CGA(Arg)	0.55	0.12	0.93
CCA(Pro)	**1.94**	2.44	1.65	CGG(Arg)	0.05	0.00	0.17
CCG(Pro)	0.43	0.12	0.74	AGA(Arg)	***2.24***	1.69	2.03
ACT(Thr)	***1.40***	1.92	1.10	AGG(Arg)	0.19	0.00	0.50
ACC(Thr)	0.19	0.03	0.35	**GGT(Gly)**	**2.17**	2.63	1.82
ACA(Thr)	***1.69***	1.81	1.49	GGC(Gly)	0.79	0.66	0.85
ACG(Thr)	0.71	0.24	1.06	GGA(Gly)	0.77	0.64	0.82
				GGG(Gly)	0.27	0.06	0.51

Table 4y: Codon usage data of *Staphylococcus lentus*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.40*	1.01	1.53	**GCT(Ala)**	**1.61**	1.82	1.42
TTC(Phe)	0.60	0.99	0.47	GCC(Ala)	0.19	0.06	0.35
TTA(Leu)	*3.73*	4.41	3.19	**GCA(Ala)**	**1.76**	1.82	1.70
TTG(Leu)	0.61	0.24	0.90	GCG(Ala)	0.43	0.31	0.53
CTT(Leu)	0.75	0.75	0.73	*TAT(Tyr)*	*1.56*	1.33	1.57
CTC(Leu)	0.16	0.02	0.30	*TAC(Tyr)*	0.44	0.67	0.43
CTA(Leu)	0.62	0.53	0.64	CAT(His)	*1.55*	1.32	1.58
CTG(Leu)	0.13	0.04	0.24	**CAC(His)**	0.45	0.68	0.42
ATT(Ile)	*1.57*	1.73	1.46	*CAA(Gln)*	*1.79*	1.93	1.60
ATC(Ile)	0.46	0.84	0.41	CAG(Gln)	0.21	0.07	0.40
ATA(Ile)	0.96	0.43	1.13	AAT(Asn)	*1.47*	1.11	1.53
GTT(Val)	*1.61*	1.85	1.43	*AAC(Asn)*	0.53	0.89	0.47
GTC(Val)	0.34	0.14	0.54	*AAA(Lys)*	*1.70*	1.83	1.51
GTA(Val)	*1.62*	1.83	1.42	AAG(Lys)	0.30	0.17	0.49
GTG(Val)	0.42	0.18	0.61	GAT(Asp)	*1.57*	1.46	1.58
TCT(Ser)	*1.59*	1.72	1.42	**GAC(Asp)**	0.43	0.54	0.42
TCC(Ser)	0.18	0.04	0.36	*GAA(Glu)*	*1.74*	1.84	1.58
TCA(Ser)	*1.93*	2.64	1.61	GAG(Glu)	0.26	0.16	0.42
TCG(Ser)	0.27	0.05	0.49	*TGT(Cys)*	*1.66*	1.77	1.50
AGT(Ser)	*1.55*	1.16	1.61	TGC(Cys)	0.34	0.23	0.50
AGC(Ser)	0.48	0.39	0.51	**CGT(Arg)**	**1.63**	3.29	1.17
CCT(Pro)	**1.54**	1.42	1.54	**CGC(Arg)**	0.35	0.43	0.42
CCC(Pro)	0.10	0.02	0.20	CGA(Arg)	0.56	0.13	0.90
CCA(Pro)	**1.97**	2.34	1.68	CGG(Arg)	0.05	0.00	0.20
CCG(Pro)	0.39	0.22	0.58	AGA(Arg)	*3.09*	2.13	2.65
ACT(Thr)	*1.33*	1.64	1.18	AGG(Arg)	0.30	0.02	0.66
ACC(Thr)	0.13	0.02	0.27	**GGT(Gly)**	**2.07**	2.52	1.65
ACA(Thr)	*2.03*	2.06	1.82	GGC(Gly)	0.57	0.52	0.66
ACG(Thr)	0.52	0.28	0.73	GGA(Gly)	**1.07**	0.85	1.18
				GGG(Gly)	0.30	0.10	0.51

Table 4z: Codon usage data of *Staphylococcus lugdunensis*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.56*	1.19	1.65	**GCT(Ala)**	**1.33**	1.64	1.18
TTC(Phe)	0.44	0.81	0.35	GCC(Ala)	0.40	0.19	0.51
TTA(Leu)	*3.34*	4.16	2.84	**GCA(Ala)**	**1.75**	1.83	1.62
TTG(Leu)	0.83	0.44	1.13	GCG(Ala)	0.52	0.34	0.70
CTT(Leu)	0.82	0.84	0.76	TAT(Tyr)	*1.62*	1.48	1.65
CTC(Leu)	0.23	0.10	0.27	*TAC(Tyr)*	0.38	0.52	0.35
CTA(Leu)	0.59	0.42	0.67	CAT(His)	*1.60*	1.40	1.62
CTG(Leu)	0.20	0.05	0.33	**CAC(His)**	0.40	0.60	0.38
ATT(Ile)	*1.81*	1.96	1.67	*CAA(Gln)*	*1.70*	1.94	1.51
ATC(Ile)	0.52	0.80	0.46	CAG(Gln)	0.30	0.06	0.49
ATA(Ile)	0.67	0.24	0.86	AAT(Asn)	*1.54*	1.31	1.60
GTT(Val)	*1.52*	1.84	1.29	*AAC(Asn)*	0.46	0.69	0.40
GTC(Val)	0.52	0.27	0.64	*AAA(Lys)*	*1.63*	1.86	1.42
GTA(Val)	*1.29*	1.49	1.18	AAG(Lys)	0.37	0.14	0.58
GTG(Val)	0.66	0.40	0.89	GAT(Asp)	*1.59*	1.48	1.59
TCT(Ser)	*1.34*	1.64	1.17	**GAC(Asp)**	0.41	0.52	0.41
TCC(Ser)	0.24	0.07	0.34	*GAA(Glu)*	*1.67*	1.83	1.51
TCA(Ser)	*1.93*	2.50	1.66	GAG(Glu)	0.33	0.17	0.49
TCG(Ser)	0.31	0.08	0.57	*TGT(Cys)*	*1.56*	1.58	1.50
AGT(Ser)	*1.59*	1.19	1.62	TGC(Cys)	0.44	0.42	0.50
AGC(Ser)	0.58	0.52	0.65	**CGT(Arg)**	**2.37**	3.58	1.78
CCT(Pro)	**1.40**	1.33	1.28	CGC(Arg)	0.84	0.63	0.95
CCC(Pro)	0.20	0.03	0.40	CGA(Arg)	0.86	0.32	1.22
CCA(Pro)	**1.99**	2.48	1.75	CGG(Arg)	0.16	0.01	0.33
CCG(Pro)	0.41	0.16	0.57	*AGA(Arg)*	*1.60*	1.45	1.36
ACT(Thr)	*1.05*	1.36	0.88	AGG(Arg)	0.17	0.01	0.37
ACC(Thr)	0.32	0.10	0.44	**GGT(Gly)**	**1.96**	2.46	1.62
ACA(Thr)	*1.93*	2.18	1.71	GGC(Gly)	0.73	0.66	0.78
ACG(Thr)	0.70	0.36	0.97	GGA(Gly)	0.98	0.79	1.07
				GGG(Gly)	0.33	0.09	0.54

Table 4aa: Codon usage data of *Staphylococcus lutrae*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.50***	1.18	1.57	**GCT(Ala)**	0.81	1.00	0.69
TTC(Phe)	0.50	0.82	0.43	GCC(Ala)	0.50	0.24	0.66
TTA(Leu)	***2.72***	3.59	2.23	**GCA(Ala)**	**1.67**	2.01	1.47
TTG(Leu)	***1.25***	0.79	1.61	GCG(Ala)	**1.02**	0.76	1.19
CTT(Leu)	0.80	0.97	0.67	TAT(Tyr)	***1.52***	1.32	1.52
CTC(Leu)	0.54	0.27	0.63	***TAC(Tyr)***	0.48	0.68	0.48
CTA(Leu)	0.37	0.29	0.38	CAT(His)	***1.53***	1.34	1.57
CTG(Leu)	0.33	0.09	0.48	**CAC(His)**	0.47	0.66	0.43
ATT(Ile)	***1.89***	1.88	1.70	***CAA(Gln)***	***1.67***	1.89	1.48
ATC(Ile)	0.70	0.96	0.70	CAG(Gln)	0.33	0.11	0.52
ATA(Ile)	0.41	0.15	0.60	AAT(Asn)	***1.43***	1.15	1.46
GTT(Val)	***1.17***	1.52	0.99	***AAC(Asn)***	0.57	0.85	0.54
GTC(Val)	0.90	1.52	0.98	***AAA(Lys)***	***1.62***	1.79	1.44
GTA(Val)	0.90	1.25	0.78	AAG(Lys)	0.38	0.21	0.56
GTG(Val)	**1.03**	0.73	1.24	GAT(Asp)	***1.50***	1.37	1.56
TCT(Ser)	***1.20***	1.51	1.01	**GAC(Asp)**	0.50	0.63	0.44
TCC(Ser)	0.37	0.14	0.43	***GAA(Glu)***	***1.58***	1.73	1.45
TCA(Ser)	***1.77***	2.47	1.50	GAG(Glu)	0.42	0.27	0.55
TCG(Ser)	0.62	0.28	0.88	***TGT(Cys)***	***1.60***	1.64	1.53
AGT(Ser)	***1.50***	1.08	1.60	TGC(Cys)	0.40	0.36	0.47
AGC(Ser)	0.54	0.53	0.58	**CGT(Arg)**	**2.62**	3.88	2.00
CCT(Pro)	**1.30**	1.31	1.15	CGC(Arg)	0.92	0.76	1.01
CCC(Pro)	0.31	0.08	0.56	CGA(Arg)	0.98	0.28	1.27
CCA(Pro)	**1.51**	2.15	1.22	CGG(Arg)	0.35	0.03	0.67
CCG(Pro)	0.88	0.46	1.08	***AGA(Arg)***	0.97	1.03	0.73
ACT(Thr)	0.73	1.08	0.58	AGG(Arg)	0.15	0.02	0.32
ACC(Thr)	0.34	0.08	0.50	**GGT(Gly)**	**1.66**	2.43	1.28
ACA(Thr)	***1.82***	2.26	1.54	GGC(Gly)	0.71	0.59	0.76
ACG(Thr)	**1.11**	0.59	1.39	GGA(Gly)	0.94	0.71	0.98
				GGG(Gly)	0.69	0.28	0.98

Table 4ab: Codon usage data of *Staphylococcus massiliensi*

Codon(Aa)	Overall (RSCU)	PHX (RSCU1.35)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.20***	0.67	1.32	**GCT(Ala)**	**1.32**	1.78	0.90
TTC(Phe)	0.80	1.33	0.68	GCC(Ala)	0.35	0.10	0.48
TTA(Leu)	***3.11***	1.34	2.22	**GCA(Ala)**	**1.37**	1.60	1.07
TTG(Leu)	0.48	0.05	0.96	GCG(Ala)	0.97	0.52	1.54
CTT(Leu)	***1.20***	1.16	1.07	TAT(Tyr)	***1.35***	1.03	1.38
CTC(Leu)	0.41	0.04	0.70	***TAC(Tyr)***	0.65	0.97	0.62
CTA(Leu)	0.62	0.38	0.69	CAT(His)	***1.34***	0.90	1.45
CTG(Leu)	0.18	0.03	0.36	**CAC(His)**	0.66	1.10	0.55
ATT(Ile)	***1.64***	1.47	1.51	***CAA(Gln)***	***1.71***	1.95	1.48
ATC(Ile)	0.88	1.49	0.83	CAG(Gln)	0.29	0.05	0.52
ATA(Ile)	0.48	0.04	0.66	AAT(Asn)	***1.23***	0.76	1.33
GTT(Val)	***1.59***	2.03	1.20	***AAC(Asn)***	0.77	1.24	0.67
GTC(Val)	0.51	0.10	0.82	***AAA(Lys)***	***1.40***	1.71	1.09
GTA(Val)	***1.26***	1.67	0.96	AAG(Lys)	0.60	0.29	0.91
GTG(Val)	0.63	0.20	1.03	GAT(Asp)	***1.37***	1.15	1.33
TCT(Ser)	***1.39***	1.79	1.04	**GAC(Asp)**	0.63	0.85	0.67
TCC(Ser)	0.24	0.03	0.40	***GAA(Glu)***	***1.60***	1.79	1.38
TCA(Ser)	***1.87***	2.64	1.37	GAG(Glu)	0.40	0.21	0.63
TCG(Ser)	0.49	0.05	1.05	***TGT(Cys)***	***1.49***	1.53	1.46
AGT(Ser)	***1.43***	0.88	1.49	TGC(Cys)	0.51	0.47	0.54
AGC(Ser)	0.58	0.61	0.64	**CGT(Arg)**	**2.88**	4.27	2.15
CCT(Pro)	**1.57**	1.50	1.35	CGC(Arg)	0.94	0.73	1.22
CCC(Pro)	0.14	0.01	0.27	CGA(Arg)	0.68	0.07	1.09
CCA(Pro)	**1.66**	2.38	1.13	CGG(Arg)	0.07	0.00	0.28
CCG(Pro)	0.63	0.12	1.24	***AGA(Arg)***	***1.24***	0.93	0.83
ACT(Thr)	0.9	1.62	0.61	AGG(Arg)	0.19	0.00	0.43
ACC(Thr)	0.26	0.02	0.51	**GGT(Gly)**	**2.31**	2.99	1.80
ACA(Thr)	***1.56***	1.93	1.05	GGC(Gly)	0.80	0.62	1.02
ACG(Thr)	**1.19**	0.43	1.83	GGA(Gly)	0.59	0.33	0.59
				GGG(Gly)	0.29	0.07	0.60

Table 4ac: Codon usage data of *Staphylococcus microti*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.38*	0.84	1.58	**GCT(Ala)**	0.73	0.90	0.55
TTC(Phe)	0.62	1.16	0.42	GCC(Ala)	0.36	0.18	0.45
TTA(Leu)	*2.74*	3.92	2.07	**GCA(Ala)**	**1.84**	2.11	1.64
TTG(Leu)	*1.14*	0.57	1.43	GCG(Ala)	**1.07**	0.81	1.36
CTT(Leu)	0.90	1.09	0.76	TAT(Tyr)	*1.41*	0.98	1.54
CTC(Leu)	0.45	0.12	0.79	*TAC(Tyr)*	0.59	1.02	0.46
CTA(Leu)	0.45	0.25	0.46	CAT(His)	*1.45*	1.01	1.52
CTG(Leu)	0.32	0.06	0.49	**CAC(His)**	0.55	0.99	0.48
ATT(Ile)	*1.97*	1.77	1.90	*CAA(Gln)*	*1.68*	1.93	1.46
ATC(Ile)	0.67	1.20	0.54	CAG(Gln)	0.32	0.07	0.54
ATA(Ile)	0.36	0.03	0.56	AAT(Asn)	*1.24*	0.75	1.41
GTT(Val)	*1.26*	1.52	1.09	*AAC(Asn)*	0.76	1.25	0.59
GTC(Val)	0.70	0.32	0.76	*AAA(Lys)*	*1.58*	1.79	1.33
GTA(Val)	*1.12*	1.62	0.95	AAG(Lys)	0.42	0.21	0.67
GTG(Val)	0.92	0.54	1.19	GAT(Asp)	*1.34*	1.12	1.40
TCT(Ser)	*1.15*	1.20	0.82	**GAC(Asp)**	0.66	0.88	0.60
TCC(Ser)	0.23	0.02	0.32	*GAA(Glu)*	*1.61*	1.76	1.34
TCA(Ser)	*1.92*	2.83	1.68	GAG(Glu)	0.39	0.24	0.66
TCG(Ser)	0.55	0.10	1.09	*TGT(Cys)*	*1.61*	1.69	1.58
AGT(Ser)	*1.43*	0.92	1.56	TGC(Cys)	0.39	0.31	0.42
AGC(Ser)	0.71	0.93	0.54	**CGT(Arg)**	**3.20**	4.47	2.50
CCT(Pro)	0.95	0.89	0.85	CGC(Arg)	**1.02**	0.79	1.10
CCC(Pro)	0.16	0.01	0.28	CGA(Arg)	0.64	0.12	0.97
CCA(Pro)	**1.97**	2.70	1.53	CGG(Arg)	0.10	0.00	0.35
CCG(Pro)	0.93	0.40	1.34	AGA(Arg)	0.92	0.62	0.79
ACT(Thr)	0.58	0.89	0.50	AGG(Arg)	0.13	0.00	0.30
ACC(Thr)	0.20	0.03	0.30	**GGT(Gly)**	**1.96**	2.75	1.45
ACA(Thr)	*2.12*	2.63	1.60	GGC(Gly)	0.91	0.68	0.95
ACG(Thr)	**1.10**	0.45	1.60	GGA(Gly)	0.62	0.42	0.75
				GGG(Gly)	0.50	0.15	0.85

Table 4ad: Codon usage data of *Staphylococcus muscae*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.30***	0.71	1.49	**GCT(Ala)**	0.82	1.03	0.65
TTC(Phe)	0.70	1.29	0.51	GCC(Ala)	0.34	0.15	0.40
TTA(Leu)	***2.59***	3.92	1.84	**GCA(Ala)**	**2.04**	2.25	1.83
TTG(Leu)	***1.20***	0.51	1.66	GCG(Ala)	0.81	0.57	1.13
CTT(Leu)	0.93	1.07	0.81	TAT(Tyr)	***1.46***	0.99	1.56
CTC(Leu)	0.47	0.14	0.59	***TAC(Tyr)***	0.54	1.01	0.44
CTA(Leu)	0.48	0.30	0.54	CAT(His)	***1.47***	1.02	1.58
CTG(Leu)	0.32	0.06	0.57	**CAC(His)**	0.53	0.98	0.42
ATT(Ile)	***1.82***	1.55	1.77	***CAA(Gln)***	***1.66***	1.93	1.44
ATC(Ile)	0.79	1.42	0.62	CAG(Gln)	0.34	0.07	0.56
ATA(Ile)	0.39	0.03	0.62	AAT(Asn)	***1.30***	0.77	1.46
GTT(Val)	***1.33***	1.63	1.03	***AAC(Asn)***	0.70	1.23	0.54
GTC(Val)	0.66	0.31	0.78	***AAA(Lys)***	***1.52***	1.77	1.27
GTA(Val)	***1.16***	1.61	0.97	AAG(Lys)	0.48	0.23	0.73
GTG(Val)	0.85	0.45	1.22	GAT(Asp)	***1.45***	1.18	1.50
TCT(Ser)	***1.22***	1.37	0.96	**GAC(Asp)**	0.55	0.82	0.50
TCC(Ser)	0.22	0.02	0.27	***GAA(Glu)***	***1.62***	1.77	1.44
TCA(Ser)	***2.11***	3.03	1.88	GAG(Glu)	0.38	0.23	0.56
TCG(Ser)	0.47	0.09	0.90	***TGT(Cys)***	***1.59***	1.63	1.58
AGT(Ser)	***1.43***	0.83	1.56	TGC(Cys)	0.41	0.37	0.42
AGC(Ser)	0.55	0.66	0.81	**CGT(Arg)**	**3.15**	4.49	2.43
CCT(Pro)	**1.01**	1.02	0.85	CGC(Arg)	0.88	0.75	0.80
CCC(Pro)	0.19	0.03	0.32	CGA(Arg)	0.83	0.13	1.38
CCA(Pro)	**2.08**	2.70	1.77	CGG(Arg)	0.13	0.00	0.36
CCG(Pro)	0.73	0.25	1.05	AGA(Arg)	0.90	0.62	0.81
ACT(Thr)	0.68	0.95	0.54	AGG(Arg)	0.11	0.01	0.21
ACC(Thr)	0.18	0.03	0.23	**GGT(Gly)**	**2.01**	2.78	1.63
ACA(Thr)	***2.29***	2.71	1.91	GGC(Gly)	0.64	0.55	0.63
ACG(Thr)	0.85	0.31	1.33	GGA(Gly)	0.89	0.54	1.03
				GGG(Gly)	0.46	0.13	0.71

Table 4ae: Codon usage data of *Staphylococcus nepalensis*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.48***	1.03	1.54	**GCT(Ala)**	**1.34**	1.52	1.14
TTC(Phe)	0.52	0.97	0.46	GCC(Ala)	0.31	0.15	0.48
TTA(Leu)	***3.53***	4.18	3.00	**GCA(Ala)**	**1.74**	1.92	1.58
TTG(Leu)	0.79	0.40	1.12	GCG(Ala)	0.61	0.41	0.80
CTT(Leu)	0.73	0.82	0.70	TAT(Tyr)	***1.55***	1.39	1.55
CTC(Leu)	0.22	0.06	0.32	***TAC(Tyr)***	0.45	0.61	0.45
CTA(Leu)	0.54	0.49	0.52	CAT(His)	***1.54***	1.21	1.60
CTG(Leu)	0.18	0.04	0.34	**CAC(His)**	0.46	0.75	0.40
ATT(Ile)	***1.70***	1.78	1.57	***CAA(Gln)***	***1.77***	1.94	1.60
ATC(Ile)	0.59	1.03	0.55	CAG(Gln)	0.23	0.06	0.40
ATA(Ile)	0.71	0.20	0.89	AAT(Asn)	***1.44***	1.16	1.48
GTT(Val)	***1.43***	1.68	1.21	***AAC(Asn)***	0.56	0.84	0.52
GTC(Val)	0.48	0.25	0.66	***AAA(Lys)***	***1.71***	1.88	1.51
GTA(Val)	***1.42***	1.66	1.28	AAG(Lys)	0.29	0.12	0.49
GTG(Val)	0.66	0.41	0.85	GAT(Asp)	***1.54***	1.41	1.56
TCT(Ser)	***1.47***	1.56	1.30	**GAC(Asp)**	0.46	0.59	0.44
TCC(Ser)	0.25	0.09	0.37	***GAA(Glu)***	***1.71***	1.85	1.55
TCA(Ser)	***1.84***	2.63	1.56	GAG(Glu)	0.29	0.15	0.45
TCG(Ser)	0.32	0.08	0.56	***TGT(Cys)***	***1.56***	1.52	1.41
AGT(Ser)	***1.52***	1.09	1.59	TGC(Cys)	0.44	0.48	0.59
AGC(Ser)	0.59	0.55	0.62	**CGT(Arg)**	**2.30**	3.87	1.79
CCT(Pro)	**1.42**	1.31	1.39	CGC(Arg)	0.64	0.64	0.68
CCC(Pro)	0.22	0.03	0.35	CGA(Arg)	0.59	0.13	0.92
CCA(Pro)	**1.88**	2.41	1.51	CGG(Arg)	0.13	0.01	0.26
CCG(Pro)	0.48	0.25	0.74	AGA(Arg)	***2.07***	1.32	1.78
ACT(Thr)	***1.15***	1.49	1.02	AGG(Arg)	0.27	0.03	0.56
ACC(Thr)	0.27	0.06	0.42	**GGT(Gly)**	**1.93**	2.44	1.61
ACA(Thr)	***1.94***	2.14	1.76	GGC(Gly)	0.87	0.79	0.91
ACG(Thr)	0.64	0.30	0.80	GGA(Gly)	0.82	0.66	0.89
				GGG(Gly)	0.38	0.11	0.58

Table 4af: Codon usage data of *Staphylococcus pasteuri*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.39*	0.96	1.40	**GCT(Ala)**	**1.37**	1.68	0.92
TTC(Phe)	0.61	1.04	0.60	GCC(Ala)	0.37	0.14	0.79
TTA(Leu)	*3.67*	4.71	2.28	**GCA(Ala)**	**1.77**	1.89	1.23
TTG(Leu)	0.63	0.19	1.05	GCG(Ala)	0.49	0.28	1.05
CTT(Leu)	0.73	0.68	0.86	TAT(Tyr)	*1.58*	1.36	1.43
CTC(Leu)	0.21	0.02	0.77	*TAC(Tyr)*	0.42	0.64	0.57
CTA(Leu)	0.63	0.38	0.53	CAT(His)	*1.59*	1.32	1.40
CTG(Leu)	0.13	0.02	0.51	**CAC(His)**	0.41	0.68	0.60
ATT(Ile)	*1.75*	1.86	1.44	*CAA(Gln)*	*1.79*	1.95	1.43
ATC(Ile)	0.61	1.00	0.65	CAG(Gln)	0.21	0.05	0.57
ATA(Ile)	0.64	0.14	0.91	AAT(Asn)	*1.53*	1.26	1.37
GTT(Val)	*1.48*	1.85	1.02	*AAC(Asn)*	0.47	0.74	0.63
GTC(Val)	0.50	0.20	0.99	*AAA(Lys)*	*1.66*	1.88	1.26
GTA(Val)	*1.45*	1.67	1.05	AAG(Lys)	0.34	0.12	0.74
GTG(Val)	0.58	0.28	0.94	*GAT(Asp)*	*1.56*	1.43	1.21
TCT(Ser)	*1.44*	1.67	1.18	GAC(Asp)	0.44	0.57	0.79
TCC(Ser)	0.23	0.05	0.50	*GAA(Glu)*	*1.70*	1.85	1.22
TCA(Ser)	*1.98*	2.60	1.28	GAG(Glu)	0.30	0.15	0.78
TCG(Ser)	0.28	0.04	0.93	*TGT(Cys)*	*1.58*	1.61	1.14
AGT(Ser)	*1.64*	1.22	1.33	TGC(Cys)	0.42	0.39	0.86
AGC(Ser)	0.43	0.43	1.00	**CGT(Arg)**	**2.22**	3.83	1.17
CCT(Pro)	**1.50**	1.43	1.10	CGC(Arg)	0.46	0.40	0.93
CCC(Pro)	0.19	0.02	0.57	CGA(Arg)	0.85	0.17	1.41
CCA(Pro)	**1.96**	2.46	1.25	CGG(Arg)	0.10	0.00	0.75
CCG(Pro)	0.36	0.09	1.08	*AGA(Arg)*	*2.18*	1.59	1.12
ACT(Thr)	*1.26*	1.73	0.92	AGG(Arg)	0.18	0.01	0.63
ACC(Thr)	0.26	0.03	0.64	**GGT(Gly)**	**2.21**	2.73	1.38
ACA(Thr)	*1.93*	2.03	1.45	GGC(Gly)	0.59	0.49	1.00
ACG(Thr)	0.56	0.21	0.99	GGA(Gly)	0.93	0.71	1.02
				GGG(Gly)	0.27	0.07	0.60

Table 4ag: Codon usage data of *Staphylococcus petrasii*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	1.29	1.00	1.45	GCT(Ala)	1.42	1.73	1.06
TTC(Phe)	0.71	1.00	0.55	GCC(Ala)	0.35	0.16	0.50
TTA(Leu)	3.71	4.92	2.85	GCA(Ala)	1.60	1.77	1.57
TTG(Leu)	0.52	0.11	0.83	GCG(Ala)	0.62	0.34	0.87
CTT(Leu)	0.88	0.67	1.17	TAT(Tyr)	1.55	1.38	1.66
CTC(Leu)	0.24	0	0.47	TAC(Tyr)	0.45	0.62	0.34
CTA(Leu)	0.55	0.26	0.55	CAT(His)	1.53	1.30	1.42
CTG(Leu)	0.10	0.03	0.13	CAC(His)	0.47	0.70	0.58
ATT(Ile)	1.88	1.93	1.76	CAA(Gln)	1.83	1.94	1.67
ATC(Ile)	0.66	1.02	0.60	CAG(Gln)	0.17	0.06	0.33
ATA(Ile)	0.46	0.06	0.64	AAT(Asn)	1.51	1.28	1.50
GTT(Val)	1.41	1.62	1.17	AAC(Asn)	0.49	0.72	0.50
GTC(Val)	0.45	0.19	0.50	AAA(Lys)	1.70	1.90	1.53
GTA(Val)	1.56	1.93	1.35	AAG(Lys)	0.30	0.10	0.47
GTG(Val)	0.58	0.26	0.98	GAT(Asp)	1.44	1.52	1.46
TCT(Ser)	1.30	1.83	1.51	GAC(Asp)	0.56	0.48	0.54
TCC(Ser)	0.17	0.02	0.26	GAA(Glu)	1.70	1.91	1.50
TCA(Ser)	2.07	2.33	1.75	GAG(Glu)	0.30	0.09	0.50
TCG(Ser)	0.24	0.02	0.70	TGT(Cys)	1.61	1.70	1.49
AGT(Ser)	1.51	1.25	1.26	TGC(Cys)	0.39	0.30	0.51
AGC(Ser)	0.72	0.55	0.52	CGT(Arg)	2.62	3.42	1.88
CCT(Pro)	1.59	1.59	1.58	CGC(Arg)	0.75	0.41	1.36
CCC(Pro)	0.15	0.06	0.32	CGA(Arg)	0.62	0.10	0.84
CCA(Pro)	1.91	2.30	1.43	CGG(Arg)	0.03	0.00	0.10
CCG(Pro)	0.35	0.06	0.67	AGA(Arg)	1.88	2.03	1.64
ACT(Thr)	1.35	2.04	0.98	AGG(Arg)	0.10	0.03	0.17
ACC(Thr)	0.23	0.03	0.51	GGT(Gly)	1.98	2.83	1.37
ACA(Thr)	1.77	1.77	1.41	GGC(Gly)	0.85	0.58	1.13
ACG(Thr)	0.65	0.15	1.09	GGA(Gly)	0.88	0.51	1.04
				GGG(Gly)	0.29	0.07	0.45

Table 4ah: Codon usage data of *Staphylococcus pettenkoferi*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.06*	0.61	1.20	**GCT(Ala)**	**1.13**	1.46	0.84
TTC(Phe)	0.94	1.39	0.80	GCC(Ala)	0.38	0.13	0.50
TTA(Leu)	*2.33*	3.50	1.69	**GCA(Ala)**	**1.40**	1.72	1.17
TTG(Leu)	0.92	0.35	1.39	GCG(Ala)	**1.10**	0.68	1.49
CTT(Leu)	0.93	1.27	0.70	TAT(Tyr)	*1.30*	0.86	1.38
CTC(Leu)	0.95	0.48	1.11	*TAC(Tyr)*	0.70	1.14	0.62
CTA(Leu)	0.55	0.36	0.54	CAT(His)	*1.30*	0.91	1.38
CTG(Leu)	0.32	0.03	0.58	**CAC(His)**	0.70	1.09	0.62
ATT(Ile)	*1.56*	1.27	1.56	*CAA(Gln)*	*1.60*	1.93	1.31
ATC(Ile)	*1.06*	1.66	0.96	CAG(Gln)	0.40	0.07	0.69
ATA(Ile)	0.39	0.07	0.47	AAT(Asn)	*1.21*	0.79	1.33
GTT(Val)	*1.07*	1.53	0.79	*AAC(Asn)*	0.79	1.21	0.67
GTC(Val)	0.84	0.36	1.02	*AAA(Lys)*	*1.42*	1.68	1.17
GTA(Val)	*1.10*	1.67	0.77	AAG(Lys)	0.58	0.32	0.83
GTG(Val)	0.99	0.44	1.41	GAT(Asp)	*1.35*	1.07	1.43
TCT(Ser)	*1.47*	1.80	1.09	**GAC(Asp)**	0.65	0.93	0.57
TCC(Ser)	0.34	0.07	0.53	*GAA(Glu)*	*1.48*	1.75	1.20
TCA(Ser)	*1.61*	2.42	1.15	GAG(Glu)	0.52	0.25	0.80
TCG(Ser)	0.42	0.04	0.93	*TGT(Cys)*	*1.51*	1.56	1.55
AGT(Ser)	*1.51*	0.96	1.64	TGC(Cys)	0.49	0.44	0.45
AGC(Ser)	0.66	0.72	0.67	**CGT(Arg)**	**2.94**	4.09	2.35
CCT(Pro)	**1.43**	1.42	1.31	CGC(Arg)	**1.12**	0.87	1.19
CCC(Pro)	0.19	0.03	0.33	CGA(Arg)	0.67	0.10	1.07
CCA(Pro)	**1.63**	2.39	1.13	CGG(Arg)	0.16	0	0.44
CCG(Pro)	0.74	0.16	1.24	*AGA(Arg)*	*1.02*	0.94	0.75
ACT(Thr)	*1.04*	1.70	0.69	AGG(Arg)	0.09	0	0.19
ACC(Thr)	0.28	0.03	0.46	**GGT(Gly)**	**2.06**	2.76	1.64
ACA(Thr)	*1.54*	1.94	1.09	GGC(Gly)	0.88	0.70	0.98
ACG(Thr)	**1.15**	0.34	1.76	GGA(Gly)	0.74	0.48	0.81
				GGG(Gly)	0.32	0.06	0.57

Table 4ai: Codon usage data of *Staphylococcus piscifermentans*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.46***	1.09	1.56	**GCT(Ala)**	**1.30**	1.61	1.10
TTC(Phe)	0.54	0.91	0.44	GCC(Ala)	0.29	0.12	0.37
TTA(Leu)	***3.43***	4.19	2.96	**GCA(Ala)**	**1.87**	1.95	1.75
TTG(Leu)	0.76	0.36	1.05	GCG(Ala)	0.53	0.33	0.78
CTT(Leu)	0.81	0.84	0.77	TAT(Tyr)	***1.58***	1.41	1.58
CTC(Leu)	0.24	0.04	0.37	***TAC(Tyr)***	0.42	0.59	0.42
CTA(Leu)	0.58	0.53	0.55	CAT(His)	***1.53***	1.28	1.58
CTG(Leu)	0.17	0.04	0.30	**CAC(His)**	0.47	0.72	0.42
ATT(Ile)	***1.69***	1.82	1.53	***CAA(Gln)***	***1.75***	1.93	1.56
ATC(Ile)	0.53	0.90	0.47	CAG(Gln)	0.25	0.07	0.44
ATA(Ile)	0.77	0.27	1.00	AAT(Asn)	***1.50***	1.21	1.55
GTT(Val)	***1.45***	1.82	1.26	***AAC(Asn)***	0.50	0.79	0.45
GTC(Val)	0.43	0.19	0.55	***AAA(Lys)***	***1.69***	1.87	1.50
GTA(Val)	***1.50***	1.67	1.31	AAG(Lys)	0.31	0.13	0.50
GTG(Val)	0.62	0.31	0.88	GAT(Asp)	***1.57***	1.44	1.58
TCT(Ser)	***1.42***	1.54	1.28	**GAC(Asp)**	0.43	0.56	0.42
TCC(Ser)	0.21	0.05	0.31	***GAA(Glu)***	***1.70***	1.82	1.55
TCA(Ser)	***1.87***	2.65	1.59	GAG(Glu)	0.30	0.18	0.45
TCG(Ser)	0.32	0.13	0.50	***TGT(Cys)***	***1.65***	1.77	1.53
AGT(Ser)	***1.68***	1.19	1.80	TGC(Cys)	0.35	0.23	0.47
AGC(Ser)	0.49	0.45	0.52	**CGT(Arg)**	**2.36**	3.81	1.64
CCT(Pro)	**1.44**	1.33	1.32	CGC(Arg)	0.61	0.55	0.65
CCC(Pro)	0.17	0.03	0.28	CGA(Arg)	0.58	0.15	0.97
CCA(Pro)	**1.97**	2.39	1.80	CGG(Arg)	0.08	0.00	0.21
CCG(Pro)	0.42	0.25	0.61	AGA(Arg)	***2.15***	1.49	2.03
ACT(Thr)	***1.18***	1.54	0.98	AGG(Arg)	0.22	0.00	0.49
ACC(Thr)	0.21	0.03	0.34	**GGT(Gly)**	**2.08**	2.44	1.75
ACA(Thr)	***1.97***	2.12	1.80	GGC(Gly)	0.76	0.76	0.79
ACG(Thr)	0.64	0.31	0.87	GGA(Gly)	0.86	0.70	0.97
				GGG(Gly)	0.30	0.10	0.49

Table 4aj: Codon usage data of *Staphylococcus pseudintermedius*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.42*	0.94	1.55	**GCT(Ala)**	0.85	1.02	0.70
TTC(Phe)	0.58	1.06	0.45	GCC(Ala)	0.39	0.20	0.48
TTA(Leu)	*2.91*	4.04	2.25	**GCA(Ala)**	**1.68**	2.00	1.47
TTG(Leu)	*1.10*	0.53	1.54	GCG(Ala)	**1.08**	0.78	1.35
CTT(Leu)	0.76	0.97	0.63	TAT(Tyr)	*1.47*	1.11	1.58
CTC(Leu)	0.57	0.19	0.79	*TAC(Tyr)*	0.53	0.89	0.42
CTA(Leu)	0.35	0.22	0.38	CAT(His)	*1.47*	1.07	1.53
CTG(Leu)	0.30	0.05	0.42	**CAC(His)**	0.53	0.93	0.47
ATT(Ile)	*1.95*	1.87	1.83	*CAA(Gln)*	*1.73*	1.92	1.56
ATC(Ile)	0.69	1.06	0.62	CAG(Gln)	0.27	0.08	0.44
ATA(Ile)	0.36	0.06	0.54	AAT(Asn)	*1.33*	0.90	1.45
GTT(Val)	*1.08*	1.46	0.54	*AAC(Asn)*	0.67	1.10	0.55
GTC(Val)	0.89	0.45	1.02	*AAA(Lys)*	*1.63*	1.81	1.44
GTA(Val)	0.96	1.39	0.77	AAG(Lys)	0.37	0.19	0.56
GTG(Val)	**1.07**	0.70	1.37	GAT(Asp)	*1.40*	1.19	1.42
TCT(Ser)	*1.18*	1.59	0.94	**GAC(Asp)**	0.60	0.81	0.58
TCC(Ser)	0.27	0.08	0.39	*GAA(Glu)*	*1.63*	1.75	1.45
TCA(Ser)	*1.84*	2.67	1.39	GAG(Glu)	0.37	0.25	0.55
TCG(Ser)	0.60	0.16	1.02	*TGT(Cys)*	*1.64*	1.72	1.56
AGT(Ser)	*1.55*	0.86	1.75	TGC(Cys)	0.36	0.28	0.44
AGC(Ser)	0.55	0.64	0.51	**CGT(Arg)**	**3.01**	4.44	2.08
CCT(Pro)	**1.20**	1.26	1.00	CGC(Arg)	0.92	0.68	1.07
CCC(Pro)	0.21	0.05	0.34	CGA(Arg)	0.73	0.13	1.17
CCA(Pro)	**1.66**	2.34	1.29	CGG(Arg)	0.22	0.00	0.61
CCG(Pro)	0.92	0.36	1.37	AGA(Arg)	0.97	0.73	0.76
ACT(Thr)	0.78	1.14	0.58	AGG(Arg)	0.15	0.01	0.30
ACC(Thr)	0.26	0.03	0.43	**GGT(Gly)**	**1.94**	2.68	1.47
ACA(Thr)	*1.90*	2.33	1.54	GGC(Gly)	0.82	0.64	0.90
ACG(Thr)	**1.06**	0.49	1.44	GGA(Gly)	0.70	0.47	0.85
				GGG(Gly)	0.54	0.21	0.79

Table 4ak: Codon usage data of *Staphylococcus pseudoxylosus*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.48*	1.11	1.60	**GCT(Ala)**	**1.37**	1.65	1.22
TTC(Phe)	0.52	0.89	0.40	GCC(Ala)	0.32	0.14	0.42
TTA(Leu)	*3.45*	4.21	3.02	**GCA(Ala)**	**1.78**	1.86	1.64
TTG(Leu)	0.81	0.39	1.16	GCG(Ala)	0.52	0.35	0.72
CTT(Leu)	0.75	0.83	0.69	TAT(Tyr)	*1.58*	1.43	1.61
CTC(Leu)	0.19	0.05	0.23	*TAC(Tyr)*	0.42	0.57	0.39
CTA(Leu)	0.63	0.48	0.66	CAT(His)	*1.53*	1.24	1.56
CTG(Leu)	0.16	0.04	0.25	**CAC(His)**	0.47	0.76	0.44
ATT(Ile)	*1.67*	1.83	1.55	*CAA(Gln)*	*1.74*	1.91	1.58
ATC(Ile)	0.52	0.89	0.41	CAG(Gln)	0.26	0.09	0.42
ATA(Ile)	0.80	0.28	1.04	AAT(Asn)	*1.49*	1.20	1.55
GTT(Val)	*1.52*	1.81	1.30	*AAC(Asn)*	0.51	0.80	0.45
GTC(Val)	0.40	0.21	0.48	*AAA(Lys)*	*1.66*	1.85	1.48
GTA(Val)	*1.50*	1.71	1.37	AAG(Lys)	0.34	0.15	0.52
GTG(Val)	0.58	0.27	0.84	GAT(Asp)	*1.56*	1.43	1.57
TCT(Ser)	*1.40*	1.58	1.16	**GAC(Asp)**	0.44	0.57	0.43
TCC(Ser)	0.24	0.07	0.38	*GAA(Glu)*	*1.67*	1.80	1.51
TCA(Ser)	*1.87*	2.57	1.63	GAG(Glu)	0.33	0.20	0.49
TCG(Ser)	0.33	0.09	0.56	*TGT(Cys)*	*1.62*	1.71	1.56
AGT(Ser)	*1.64*	1.17	1.76	TGC(Cys)	0.38	0.29	0.44
AGC(Ser)	0.51	0.53	0.51	**CGT(Arg)**	**2.25**	3.84	1.48
CCT(Pro)	**1.46**	1.40	1.45	CGC(Arg)	0.54	0.49	0.57
CCC(Pro)	0.18	0.03	0.28	CGA(Arg)	0.56	0.11	0.87
CCA(Pro)	**1.96**	2.38	1.67	CGG(Arg)	0.09	0.00	0.23
CCG(Pro)	0.40	0.19	0.60	AGA(Arg)	*2.26*	1.55	2.10
ACT(Thr)	*1.25*	1.58	1.07	AGG(Arg)	0.31	0.01	0.75
ACC(Thr)	0.24	0.05	0.39	**GGT(Gly)**	**2.08**	2.56	1.79
ACA(Thr)	*1.92*	2.05	1.76	GGC(Gly)	0.68	0.64	0.67
ACG(Thr)	0.59	0.32	0.78	GGA(Gly)	0.90	0.69	0.99
				GGG(Gly)	0.34	0.10	0.55

Table 4al: Codon usage data of *Staphylococcus rostri*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.30*	0.73	1.52	**GCT(Ala)**	0.69	0.98	0.55
TTC(Phe)	0.70	1.27	0.48	GCC(Ala)	0.39	0.17	0.51
TTA(Leu)	*2.69*	4.02	2.02	**GCA(Ala)**	**1.80**	2.10	1.58
TTG(Leu)	*1.19*	0.53	1.55	GCG(Ala)	**1.12**	0.76	1.36
CTT(Leu)	0.84	1.08	0.69	TAT(Tyr)	*1.40*	0.89	1.55
CTC(Leu)	0.45	0.09	0.62	*TAC(Tyr)*	0.60	1.11	0.45
CTA(Leu)	0.48	0.23	0.52	CAT(His)	*1.38*	0.89	1.50
CTG(Leu)	0.35	0.04	0.58	**CAC(His)**	0.62	1.11	0.50
ATT(Ile)	*1.93*	1.54	1.83	*CAA(Gln)*	*1.64*	1.94	1.38
ATC(Ile)	0.76	1.44	0.58	CAG(Gln)	0.36	0.06	0.62
ATA(Ile)	0.31	0.02	0.59	AAT(Asn)	*1.19*	0.68	1.39
GTT(Val)	*1.24*	1.61	1.00	*AAC(Asn)*	0.81	1.32	0.61
GTC(Val)	0.72	0.31	0.83	*AAA(Lys)*	*1.52*	1.78	1.27
GTA(Val)	*1.13*	1.61	0.93	AAG(Lys)	0.48	0.22	0.73
GTG(Val)	0.91	0.47	1.24	GAT(Asp)	*1.31*	1.08	1.37
TCT(Ser)	*1.10*	1.40	0.91	**GAC(Asp)**	0.69	0.92	0.63
TCC(Ser)	0.23	0.03	0.36	*GAA(Glu)*	*1.59*	1.77	1.38
TCA(Ser)	*1.96*	2.89	1.47	GAG(Glu)	0.41	0.23	0.62
TCG(Ser)	0.56	0.10	1.06	*TGT(Cys)*	*1.56*	1.62	1.46
AGT(Ser)	*1.43*	0.78	1.60	TGC(Cys)	0.44	0.38	0.54
AGC(Ser)	0.71	0.80	0.61	**CGT(Arg)**	**3.28**	4.48	2.58
CCT(Pro)	0.93	0.96	0.77	CGC(Arg)	**1.13**	0.82	1.26
CCC(Pro)	0.16	0.01	0.32	CGA(Arg)	0.59	0.10	0.97
CCA(Pro)	**2.05**	2.76	1.53	CGG(Arg)	0.10	0.00	0.33
CCG(Pro)	0.86	0.27	1.38	AGA(Arg)	0.80	0.61	0.64
ACT(Thr)	0.56	0.93	0.41	AGG(Arg)	0.11	0.00	0.21
ACC(Thr)	0.21	0.02	0.33	**GGT(Gly)**	**2.03**	2.88	1.56
ACA(Thr)	*2.11*	2.66	1.56	GGC(Gly)	0.92	0.64	1.00
ACG(Thr)	*1.12*	0.39	1.70	GGA(Gly)	0.62	0.36	0.70
				GGG(Gly)	0.43	0.13	0.74

Table 4am: Codon usage data of *Staphylococcus saccharolyticus*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.43***	0.96	1.58	**GCT(Ala)**	**1.56**	1.87	1.37
TTC(Phe)	0.57	1.04	0.42	GCC(Ala)	0.29	0.09	0.41
TTA(Leu)	***3.88***	4.76	3.34	**GCA(Ala)**	**1.78**	1.82	1.71
TTG(Leu)	0.56	0.14	0.91	GCG(Ala)	0.38	0.22	0.52
CTT(Leu)	0.75	0.69	0.72	TAT(Tyr)	***1.61***	1.43	1.65
CTC(Leu)	0.14	0.01	0.24	***TAC(Tyr)***	0.39	0.57	0.35
CTA(Leu)	0.59	0.39	0.64	CAT(His)	***1.60***	1.31	1.65
CTG(Leu)	0.09	0.01	0.15	**CAC(His)**	0.40	0.69	0.35
ATT(Ile)	***1.80***	1.89	1.57	***CAA(Gln)***	***1.83***	1.96	1.68
ATC(Ile)	0.50	0.95	0.39	CAG(Gln)	0.17	0.04	0.32
ATA(Ile)	0.70	0.16	1.04	AAT(Asn)	***1.55***	1.26	1.57
GTT(Val)	***1.65***	1.90	1.33	***AAC(Asn)***	0.45	0.74	0.43
GTC(Val)	0.38	0.14	0.56	***AAA(Lys)***	***1.73***	1.92	1.54
GTA(Val)	***1.53***	1.76	1.42	AAG(Lys)	0.27	0.08	0.46
GTG(Val)	0.44	0.21	0.68	GAT(Asp)	***1.58***	1.42	1.56
TCT(Ser)	***1.51***	1.70	1.31	**GAC(Asp)**	0.42	0.58	0.44
TCC(Ser)	0.17	0.02	0.23	***GAA(Glu)***	***1.77***	1.89	1.58
TCA(Ser)	***2.02***	2.67	1.70	GAG(Glu)	0.23	0.11	0.42
TCG(Ser)	0.20	0.01	0.44	***TGT(Cys)***	***1.65***	1.71	1.65
AGT(Ser)	***1.68***	1.20	1.81	TGC(Cys)	0.35	0.29	0.35
AGC(Ser)	0.41	0.40	0.51	**CGT(Arg)**	**2.29**	3.93	1.55
CCT(Pro)	**1.58**	1.45	1.55	CGC(Arg)	0.37	0.35	0.44
CCC(Pro)	0.14	0.02	0.32	CGA(Arg)	0.71	0.12	1.11
CCA(Pro)	**2.01**	2.48	1.69	CGG(Arg)	0.04	0.00	0.15
CCG(Pro)	0.26	0.05	0.44	AGA(Arg)	***2.45***	1.59	2.31
ACT(Thr)	***1.51***	1.95	1.18	AGG(Arg)	0.14	0.01	0.44
ACC(Thr)	0.17	0.02	0.30	**GGT(Gly)**	**2.19**	2.68	1.77
ACA(Thr)	***1.90***	1.91	1.78	GGC(Gly)	0.54	0.46	0.60
ACG(Thr)	0.42	0.12	0.74	GGA(Gly)	**1.06**	0.81	1.23
				GGG(Gly)	0.22	0.06	0.39

Table 4an: Codon usage data of *Staphylococcus saprophyticus*

sCodon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.54***	1.19	1.61	**GCT(Ala)**	**1.34**	1.53	1.16
TTC(Phe)	0.46	0.81	0.39	GCC(Ala)	0.28	0.10	0.45
TTA(Leu)	***3.48***	4.20	3.00	**GCA(Ala)**	**1.89**	2.05	1.75
TTG(Leu)	0.84	0.43	1.15	GCG(Ala)	0.49	0.31	0.64
CTT(Leu)	0.74	0.78	0.74	TAT(Tyr)	***1.62***	1.56	1.58
CTC(Leu)	0.20	0.05	0.28	***TAC(Tyr)***	0.38	0.44	0.42
CTA(Leu)	0.57	0.49	0.56	CAT(His)	***1.58***	1.35	1.60
CTG(Leu)	0.18	0.05	0.26	**CAC(His)**	0.42	0.65	0.40
ATT(Ile)	***1.68***	1.83	1.53	***CAA(Gln)***	***1.75***	1.93	1.54
ATC(Ile)	0.53	0.81	0.54	CAG(Gln)	0.25	0.07	0.46
ATA(Ile)	0.79	0.36	0.94	AAT(Asn)	***1.54***	1.31	1.51
GTT(Val)	***1.51***	1.75	1.29	***AAC(Asn)***	0.46	0.69	0.49
GTC(Val)	0.44	0.24	0.54	***AAA(Lys)***	***1.70***	1.85	1.49
GTA(Val)	***1.44***	1.67	1.29	AAG(Lys)	0.30	0.15	0.51
GTG(Val)	0.61	0.33	0.88	GAT(Asp)	***1.60***	1.46	1.54
TCT(Ser)	***1.44***	1.57	1.23	**GAC(Asp)**	0.40	0.54	0.46
TCC(Ser)	0.24	0.07	0.34	***GAA(Glu)***	***1.72***	1.84	1.60
TCA(Ser)	***1.91***	2.71	1.65	GAG(Glu)	0.28	0.16	0.40
TCG(Ser)	0.30	0.06	0.59	***TGT(Cys)***	***1.62***	1.72	1.58
AGT(Ser)	***1.63***	1.17	1.62	TGC(Cys)	0.38	0.28	0.42
AGC(Ser)	0.48	0.42	0.58	**CGT(Arg)**	**2.19**	3.73	1.75
CCT(Pro)	**1.47**	1.37	1.45	CGC(Arg)	0.55	0.45	0.63
CCC(Pro)	0.19	0.03	0.38	CGA(Arg)	0.63	0.18	0.90
CCA(Pro)	**1.96**	2.40	1.63	CGG(Arg)	0.11	0.00	0.27
CCG(Pro)	0.39	0.20	0.54	AGA(Arg)	***2.26***	1.62	1.85
ACT(Thr)	***1.19***	1.47	1.07	AGG(Arg)	0.26	0.02	0.60
ACC(Thr)	0.26	0.06	0.41	**GGT(Gly)**	**2.13**	2.56	1.74
ACA(Thr)	***1.96***	2.14	1.79	GGC(Gly)	0.65	0.56	0.75
ACG(Thr)	0.59	0.32	0.74	GGA(Gly)	0.93	0.79	1.02
				GGG(Gly)	0.29	0.08	0.49

Table 4ao: Codon usage data of *Staphylococcus schleiferi*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.49*	1.09	1.62	**GCT(Ala)**	**1.16**	1.42	0.92
TTC(Phe)	0.51	0.91	0.38	GCC(Ala)	0.34	0.14	0.52
TTA(Leu)	*3.32*	4.32	2.66	**GCA(Ala)**	**1.79**	1.96	1.56
TTG(Leu)	0.90	0.45	1.15	GCG(Ala)	0.71	0.48	1.01
CTT(Leu)	0.82	0.79	0.81	TAT(Tyr)	*1.55*	1.30	1.58
CTC(Leu)	0.28	0.04	0.55	*TAC(Tyr)*	0.45	0.70	0.42
CTA(Leu)	0.49	0.36	0.50	CAT(His)	*1.57*	1.29	1.59
CTG(Leu)	0.19	0.04	0.33	**CAC(His)**	0.43	0.71	0.41
ATT(Ile)	*1.86*	1.97	1.76	*CAA(Gln)*	*1.74*	1.93	1.57
ATC(Ile)	0.59	0.88	0.63	CAG(Gln)	0.26	0.07	0.43
ATA(Ile)	0.55	0.15	0.61	AAT(Asn)	*1.49*	1.21	1.53
GTT(Val)	*1.53*	1.89	1.21	*AAC(Asn)*	0.51	0.79	0.47
GTC(Val)	0.59	0.22	0.89	*AAA(Lys)*	*1.64*	1.82	1.50
GTA(Val)	*1.21*	1.54	0.90	AAG(Lys)	0.36	0.18	0.50
GTG(Val)	0.68	0.36	1.01	GAT(Asp)	*1.55*	1.40	1.54
TCT(Ser)	*1.34*	1.53	1.24	**GAC(Asp)**	0.45	0.60	0.46
TCC(Ser)	0.21	0.03	0.39	*GAA(Glu)*	*1.65*	1.80	1.50
TCA(Ser)	*1.93*	2.57	1.52	GAG(Glu)	0.35	0.20	0.50
TCG(Ser)	0.42	0.09	0.71	*TGT(Cys)*	*1.61*	1.69	1.61
AGT(Ser)	*1.58*	1.17	1.62	TGC(Cys)	0.39	0.31	0.39
AGC(Ser)	0.52	0.59	0.52	**CGT(Arg)**	**2.45**	3.85	1.93
CCT(Pro)	**1.43**	1.41	1.36	CGC(Arg)	0.71	0.55	0.95
CCC(Pro)	0.18	0.02	0.45	CGA(Arg)	0.86	0.26	1.16
CCA(Pro)	**1.83**	2.35	1.34	CGG(Arg)	0.15	0.00	0.46
CCG(Pro)	0.57	0.22	0.84	*AGA(Arg)*	*1.64*	1.32	1.17
ACT(Thr)	*1.04*	1.46	0.74	AGG(Arg)	0.19	0.01	0.33
ACC(Thr)	0.23	0.03	0.43	**GGT(Gly)**	**2.03**	2.70	1.49
ACA(Thr)	*1.99*	2.19	1.69	GGC(Gly)	0.70	0.54	0.87
ACG(Thr)	0.75	0.32	1.14	GGA(Gly)	0.89	0.65	0.94
				GGG(Gly)	0.38	0.10	0.70

Table 4ap: Codon usage data of *Staphylococcus sciuri*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.39***	0.98	1.55	**GCT(Ala)**	**1.32**	1.64	1.13
TTC(Phe)	0.61	1.02	0.45	GCC(Ala)	0.24	0.09	0.47
TTA(Leu)	***3.68***	4.51	3.11	**GCA(Ala)**	**1.93**	1.96	1.69
TTG(Leu)	0.63	0.24	1.02	GCG(Ala)	0.51	0.31	0.71
CTT(Leu)	0.81	0.81	0.70	TAT(Tyr)	***1.57***	1.28	1.62
CTC(Leu)	0.22	0.03	0.36	***TAC(Tyr)***	0.43	0.72	0.38
CTA(Leu)	0.52	0.39	0.57	CAT(His)	***1.59***	1.32	1.63
CTG(Leu)	0.14	0.02	0.24	**CAC(His)**	0.41	0.68	0.37
ATT(Ile)	***1.68***	1.78	1.47	***CAA(Gln)***	***1.80***	1.96	1.65
ATC(Ile)	0.58	0.99	0.51	CAG(Gln)	0.20	0.04	0.35
ATA(Ile)	0.74	0.23	1.02	AAT(Asn)	***1.49***	1.12	1.59
GTT(Val)	***1.59***	1.90	1.26	***AAC(Asn)***	0.51	0.88	0.41
GTC(Val)	0.47	0.19	0.61	***AAA(Lys)***	***1.66***	1.84	1.48
GTA(Val)	***1.48***	1.71	1.38	AAG(Lys)	0.34	0.16	0.52
GTG(Val)	0.46	0.20	0.75	GAT(Asp)	***1.57***	1.43	1.54
TCT(Ser)	***1.51***	1.74	1.30	**GAC(Asp)**	0.43	0.57	0.46
TCC(Ser)	0.16	0.02	0.32	***GAA(Glu)***	***1.75***	1.85	1.56
TCA(Ser)	***2.04***	2.67	1.67	GAG(Glu)	0.25	0.15	0.44
TCG(Ser)	0.26	0.04	0.53	***TGT(Cys)***	***1.68***	1.80	1.58
AGT(Ser)	***1.61***	1.15	1.69	TGC(Cys)	0.32	0.20	0.42
AGC(Ser)	0.42	0.38	0.49	**CGT(Arg)**	**1.74**	3.33	1.20
CCT(Pro)	**1.38**	1.30	1.25	CGC(Arg)	0.37	0.37	0.51
CCC(Pro)	0.09	0.01	0.20	CGA(Arg)	0.61	0.11	1.17
CCA(Pro)	**2.12**	2.58	1.82	CGG(Arg)	0.05	0.00	0.16
CCG(Pro)	0.40	0.11	0.73	***AGA(Arg)***	***3.01***	2.16	2.49
ACT(Thr)	***1.08***	1.45	0.89	AGG(Arg)	0.23	0.02	0.47
ACC(Thr)	0.13	0.01	0.26	**GGT(Gly)**	**2.22**	2.73	1.79
ACA(Thr)	***2.14***	2.27	1.90	GGC(Gly)	0.55	0.44	0.65
ACG(Thr)	0.65	0.27	0.95	GGA(Gly)	0.98	0.76	1.06
				GGG(Gly)	0.24	0.07	0.49

Table 4aq: Codon usage data of *Staphylococcus simiae*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.51***	1.16	1.61	**GCT(Ala)**	**1.52**	1.79	1.26
TTC(Phe)	0.49	0.84	0.39	GCC(Ala)	0.33	0.13	0.46
TTA(Leu)	***3.87***	4.63	3.17	**GCA(Ala)**	**1.76**	1.85	1.72
TTG(Leu)	0.69	0.30	0.99	GCG(Ala)	0.39	0.24	0.56
CTT(Leu)	0.61	0.65	0.63	TAT(Tyr)	***1.65***	1.52	1.63
CTC(Leu)	0.12	0.01	0.24	***TAC(Tyr)***	0.35	0.48	0.37
CTA(Leu)	0.57	0.40	0.70	CAT(His)	***1.65***	1.43	1.68
CTG(Leu)	0.13	0.02	0.27	**CAC(His)**	0.35	0.57	0.32
ATT(Ile)	***1.85***	2.03	1.60	***CAA(Gln)***	***1.80***	1.95	1.54
ATC(Ile)	0.48	0.77	0.43	CAG(Gln)	0.20	0.05	0.46
ATA(Ile)	0.67	0.20	0.97	AAT(Asn)	***1.57***	1.38	1.52
GTT(Val)	***1.64***	1.90	1.34	***AAC(Asn)***	0.43	0.62	0.48
GTC(Val)	0.51	0.24	0.69	***AAA(Lys)***	***1.70***	1.88	1.48
GTA(Val)	***1.35***	1.58	1.18	AAG(Lys)	0.30	0.12	0.52
GTG(Val)	0.51	0.28	0.79	GAT(Asp)	***1.61***	1.50	1.55
TCT(Ser)	***1.35***	1.45	1.12	**GAC(Asp)**	0.39	0.50	0.45
TCC(Ser)	0.18	0.05	0.28	***GAA(Glu)***	***1.73***	1.86	1.53
TCA(Ser)	***1.98***	2.67	1.74	GAG(Glu)	0.27	0.14	0.47
TCG(Ser)	0.24	0.04	0.48	***TGT(Cys)***	***1.66***	1.72	1.61
AGT(Ser)	***1.87***	1.87	1.91	TGC(Cys)	0.34	0.28	0.39
AGC(Ser)	0.38	0.39	0.47	**CGT(Arg)**	**2.45**	3.90	1.50
CCT(Pro)	**1.42**	1.28	1.28	CGC(Arg)	0.48	0.35	0.59
CCC(Pro)	0.13	0.02	0.28	CGA(Arg)	0.65	0.17	1.10
CCA(Pro)	**2.11**	2.56	1.78	CGG(Arg)	0.05	0.00	0.12
CCG(Pro)	0.33	0.14	0.66	AGA(Arg)	***2.19***	1.58	2.17
ACT(Thr)	***1.35***	1.72	1.15	AGG(Arg)	0.19	0.00	0.51
ACC(Thr)	0.21	0.04	0.35	**GGT(Gly)**	**2.27**	2.69	1.84
ACA(Thr)	***1.91***	2.00	1.77	GGC(Gly)	0.55	0.46	0.64
ACG(Thr)	0.54	0.23	0.72	GGA(Gly)	0.95	0.78	1.09
				GGG(Gly)	0.23	0.07	0.43

Table 4ar: Codon usage data of *Staphylococcus simulans*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.07***	0.54	1.37	**GCT(Ala)**	**1.02**	1.39	0.89
TTC(Phe)	0.93	1.46	0.63	GCC(Ala)	0.30	0.10	0.42
TTA(Leu)	***3.43***	4.35	2.81	**GCA(Ala)**	**1.81**	2.06	1.51
TTG(Leu)	0.92	0.47	1.31	GCG(Ala)	0.87	0.45	1.18
CTT(Leu)	0.71	0.8	0.62	TAT(Tyr)	***1.48***	1.12	1.60
CTC(Leu)	0.25	0.03	0.37	***TAC(Tyr)***	0.52	0.88	0.40
CTA(Leu)	0.40	0.31	0.37	CAT(His)	***1.47***	1.00	1.58
CTG(Leu)	0.30	0.03	0.51	**CAC(His)**	0.53	1.00	0.42
ATT(Ile)	***1.74***	1.36	1.65	***CAA(Gln)***	***1.79***	1.96	1.53
ATC(Ile)	0.95	1.62	0.72	CAG(Gln)	0.21	0.04	0.47
ATA(Ile)	0.31	0.03	0.62	AAT(Asn)	***1.25***	0.77	1.46
GTT(Val)	***1.22***	1.69	1.02	***AAC(Asn)***	0.75	1.23	0.54
GTC(Val)	0.78	0.31	0.92	***AAA(Lys)***	***1.70***	1.92	1.44
GTA(Val)	***1.26***	1.67	1.03	AAG(Lys)	0.30	0.08	0.56
GTG(Val)	0.75	0.33	1.03	GAT(Asp)	***1.47***	1.21	1.54
TCT(Ser)	***1.46***	1.61	1.31	**GAC(Asp)**	0.53	0.79	0.46
TCC(Ser)	0.18	0.03	0.29	***GAA(Glu)***	***1.73***	1.85	1.56
TCA(Ser)	***2.01***	2.73	1.56	GAG(Glu)	0.27	0.15	0.44
TCG(Ser)	0.25	0.03	0.53	TGT(Cys)	***1.47***	1.40	1.56
AGT(Ser)	***1.42***	0.83	1.72	**TGC(Cys)**	0.53	0.60	0.44
AGC(Ser)	0.68	0.77	0.59	***CGT(Arg)***	**3.10**	4.27	2.13
CCT(Pro)	**1.27**	1.29	1.18	CGC(Arg)	0.82	0.66	0.80
CCC(Pro)	0.12	0.02	0.25	CGA(Arg)	0.46	0.07	0.81
CCA(Pro)	**1.76**	2.39	1.34	CGG(Arg)	0.06	0.00	0.23
CCG(Pro)	0.85	0.3	1.24	AGA(Arg)	***1.47***	1.00	1.74
ACT(Thr)	***1.02***	1.58	0.90	AGG(Arg)	0.10	0.00	0.29
ACC(Thr)	0.23	0.02	0.47	**GGT(Gly)**	**2.08**	2.60	1.68
ACA(Thr)	***2.16***	2.23	1.69	GGC(Gly)	0.90	0.81	0.88
ACG(Thr)	0.60	0.17	0.95	GGA(Gly)	0.75	0.52	0.93
				GGG(Gly)	0.27	0.07	0.51

Table 4as: Codon usage data of *Staphylococcus succinus*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.48***	1.16	1.16	**GCT(Ala)**	**1.29**	1.66	0.48
TTC(Phe)	0.52	0.84	0.84	GCC(Ala)	0.41	0.17	1.22
TTA(Leu)	***3.21***	4.13	0.75	**GCA(Ala)**	**1.66**	1.77	0.77
TTG(Leu)	0.79	0.42	0.76	GCG(Ala)	0.63	0.40	1.53
CTT(Leu)	0.77	0.84	0.61	***TAT(Tyr)***	***1.57***	1.48	1.13
CTC(Leu)	0.26	0.09	0.63	TAC(Tyr)	0.43	0.52	0.87
CTA(Leu)	0.54	0.45	0.19	***CAT(His)***	***1.52***	1.37	1.18
CTG(Leu)	0.43	0.07	3.05	CAC(His)	0.48	0.63	0.82
ATT(Ile)	***1.67***	1.83	1.59	***CAA(Gln)***	***1.65***	1.90	0.65
ATC(Ile)	0.55	0.83	1.24	CAG(Gln)	0.35	0.10	1.35
ATA(Ile)	0.78	0.34	0.16	***AAT(Asn)***	***1.5***	1.34	0.88
GTT(Val)	***1.38***	1.74	0.82	AAC(Asn)	0.5	0.66	1.12
GTC(Val)	0.48	0.24	0.94	***AAA(Lys)***	***1.64***	1.82	1.5
GTA(Val)	***1.45***	1.67	0.49	AAG(Lys)	0.36	0.18	0.5
GTG(Val)	0.7	0.34	1.75	***GAT(Asp)***	***1.58***	1.51	1.27
TCT(Ser)	***1.42***	1.49	0.68	GAC(Asp)	0.42	0.49	0.73
TCC(Ser)	0.3	0.1	0.87	***GAA(Glu)***	***1.65***	1.79	1.26
TCA(Ser)	***1.71***	2.51	0.65	GAG(Glu)	0.35	0.21	0.74
TCG(Ser)	0.38	0.12	1.17	***TGT(Cys)***	***1.51***	1.7	0.82
AGT(Ser)	***1.62***	1.3	0.87	TGC(Cys)	0.49	0.3	1.18
AGC(Ser)	0.58	0.47	1.76	**CGT(Arg)**	**2.28**	3.68	1.89
CCT(Pro)	**1.4**	1.4	0.55	CGC(Arg)	0.8	0.57	2.56
CCC(Pro)	0.25	0.06	0.59	CGA(Arg)	0.58	0.17	0.51
CCA(Pro)	**1.75**	2.3	0.65	CGG(Arg)	0.16	0.01	0.79
CCG(Pro)	0.6	0.23	2.21	***AGA(Arg)***	***1.91***	1.56	0.13
ACT(Thr)	***1.16***	1.56	0.5	AGG(Arg)	0.26	0.01	0.12
ACC(Thr)	0.38	0.08	1.82	**GGT(Gly)**	**1.92**	2.45	1.12
ACA(Thr)	***1.81***	2.04	0.44	GGC(Gly)	0.78	0.67	1.66
ACG(Thr)	0.64	0.32	1.23	**GGA(Gly)**	0.91	0.74	0.46
				GGG(Gly)	0.39	0.14	0.76

Table 4at: Codon usage data of *Staphylococcus vitulinus*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	*1.41*	1.03	1.48	**GCT(Ala)**	**1.46**	1.76	1.2
TTC(Phe)	0.59	0.97	0.52	GCC(Ala)	0.23	0.08	0.37
TTA(Leu)	*3.59*	4.37	3.07	**GCA(Ala)**	**1.81**	1.84	1.69
TTG(Leu)	0.7	0.3	0.99	GCG(Ala)	0.50	0.32	0.74
CTT(Leu)	0.79	0.82	0.7	TAT(Tyr)	*1.58*	1.37	1.56
CTC(Leu)	0.23	0.03	0.42	*TAC(Tyr)*	0.42	0.63	0.44
CTA(Leu)	0.53	0.45	0.55	CAT(His)	*1.58*	1.38	1.67
CTG(Leu)	0.16	0.04	0.26	**CAC(His)**	0.42	0.62	0.33
ATT(Ile)	*1.59*	1.74	1.42	*CAA(Gln)*	*1.79*	1.94	1.62
ATC(Ile)	0.53	0.87	0.53	CAG(Gln)	0.21	0.06	0.38
ATA(Ile)	0.88	0.39	1.05	AAT(Asn)	*1.46*	1.16	1.5
GTT(Val)	*1.6*	1.86	1.36	*AAC(Asn)*	0.54	0.84	0.5
GTC(Val)	0.43	0.18	0.6	*AAA(Lys)*	*1.68*	1.86	1.47
GTA(Val)	*1.48*	1.74	1.31	AAG(Lys)	0.32	0.14	0.53
GTG(Val)	0.5	0.22	0.73	GAT(Asp)	*1.58*	1.46	1.55
TCT(Ser)	*1.51*	1.71	1.18	**GAC(Asp)**	0.42	0.54	0.45
TCC(Ser)	0.19	0.02	0.34	*GAA(Glu)*	*1.73*	1.84	1.63
TCA(Ser)	*1.95*	2.72	1.75	GAG(Glu)	0.27	0.16	0.37
TCG(Ser)	0.30	0.07	0.6	*TGT(Cys)*	*1.63*	1.79	1.47
AGT(Ser)	*1.55*	1.1	1.47	TGC(Cys)	0.37	0.21	0.53
AGC(Ser)	0.5	0.38	0.67	**CGT(Arg)**	**1.59**	3.27	1.09
CCT(Pro)	**1.52**	1.44	1.52	CGC(Arg)	0.38	0.45	0.5
CCC(Pro)	0.11	0.01	0.28	CGA(Arg)	0.58	0.15	1
CCA(Pro)	**1.95**	2.38	1.6	CGG(Arg)	0.05	0	0.27
CCG(Pro)	0.42	0.17	0.6	AGA(Arg)	*3.12*	2.1	2.55
ACT(Thr)	*1.21*	1.6	0.99	AGG(Arg)	0.27	0.02	0.59
ACC(Thr)	0.14	0.02	0.3	**GGT(Gly)**	**2.07**	2.54	1.65
ACA(Thr)	*1.96*	2.03	1.78	GGC(Gly)	0.6	0.51	0.68
ACG(Thr)	0.68	0.34	0.92	GGA(Gly)	**1.07**	0.88	1.18
				GGG(Gly)	0.27	0.07	0.49

Table 4au: Codon usage data of *Staphylococcus warneri*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	1.39	0.91	1.56	GCT(Ala)	1.36	1.68	1.06
TTC(Phe)	0.61	1.09	0.44	GCC(Ala)	0.36	0.15	0.49
TTA(Leu)	3.75	4.62	3.16	GCA(Ala)	1.81	1.87	1.72
TTG(Leu)	0.63	0.21	1.01	GCG(Ala)	0.48	0.30	0.74
CTT(Leu)	0.7	0.71	0.63	TAT(Tyr)	1.59	1.35	1.58
CTC(Leu)	0.2	0.02	0.32	TAC(Tyr)	0.41	0.65	0.42
CTA(Leu)	0.6	0.42	0.67	CAT(His)	1.59	1.31	1.59
CTG(Leu)	0.12	0.02	0.22	CAC(His)	0.41	0.69	0.41
ATT(Ile)	1.75	1.87	1.53	CAA(Gln)	1.81	1.96	1.66
ATC(Ile)	0.62	1.01	0.52	CAG(Gln)	0.19	0.04	0.34
ATA(Ile)	0.63	0.11	0.95	AAT(Asn)	1.52	1.27	1.55
GTT(Val)	1.49	1.86	1.16	AAC(Asn)	0.48	0.73	0.45
GTC(Val)	0.49	0.17	0.67	AAA(Lys)	1.66	1.87	1.44
GTA(Val)	1.44	1.68	1.3	AAG(Lys)	0.34	0.13	0.56
GTG(Val)	0.59	0.29	0.87	GAT(Asp)	1.56	1.42	1.51
TCT(Ser)	1.41	1.69	1.23	GAC(Asp)	0.44	0.58	0.49
TCC(Ser)	0.22	0.05	0.32	GAA(Glu)	1.71	1.85	1.52
TCA(Ser)	2.04	2.64	1.73	GAG(Glu)	0.29	0.15	0.48
TCG(Ser)	0.25	0.02	0.55	TGT(Cys)	1.64	1.62	1.6
AGT(Ser)	1.66	1.18	1.73	TGC(Cys)	0.36	0.38	0.4
AGC(Ser)	0.42	0.41	0.44	CGT(Arg)	2.31	3.78	1.64
CCT(Pro)	1.48	1.41	1.37	CGC(Arg)	0.42	0.42	0.44
CCC(Pro)	0.17	0.03	0.33	CGA(Arg)	0.79	0.17	1.26
CCA(Pro)	2.00	2.42	1.77	CGG(Arg)	0.05	0	0.18
CCG(Pro)	0.34	0.14	0.53	AGA(Arg)	2.26	1.62	1.99
ACT(Thr)	1.23	1.77	0.90	AGG(Arg)	0.18	0	0.49
ACC(Thr)	0.25	0.03	0.45	GGT(Gly)	2.24	2.74	1.83
ACA(Thr)	1.97	1.99	1.79	GGC(Gly)	0.59	0.49	0.66
ACG(Thr)	0.55	0.21	0.86	GGA(Gly)	0.91	0.71	1.03
				GGG(Gly)	0.26	0.06	0.48

Table 4av: Codon usage data of *Staphylococcus xylosus*

Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)	Codon(Aa)	Overall (RSCU)	PHX (RSCU)	PLX (RSCU)
TTT(Phe)	***1.49***	1.12	1.59	**GCT(Ala)**	**1.39**	1.67	1.19
TTC(Phe)	0.51	0.88	0.41	GCC(Ala)	0.31	0.13	0.42
TTA(Leu)	***3.46***	4.25	2.99	**GCA(Ala)**	**1.79**	1.85	1.67
TTG(Leu)	0.8	0.36	1.14	GCG(Ala)	0.51	0.35	0.72
CTT(Leu)	0.74	0.82	0.65	TAT(Tyr)	***1.58***	1.45	1.6
CTC(Leu)	0.19	0.05	0.24	***TAC(Tyr)***	0.42	0.55	0.4
CTA(Leu)	0.64	0.5	0.72	CAT(His)	***1.54***	1.27	1.62
CTG(Leu)	0.16	0.03	0.26	**CAC(His)**	0.46	0.73	0.38
ATT(Ile)	***1.67***	1.8	1.54	***CAA(Gln)***	***1.74***	1.9	1.6
ATC(Ile)	0.5	0.85	0.42	CAG(Gln)	0.26	0.1	0.4
ATA(Ile)	0.83	0.35	1.04	AAT(Asn)	***1.52***	1.27	1.6
GTT(Val)	***1.53***	1.86	1.33	***AAC(Asn)***	0.48	0.73	0.4
GTC(Val)	0.39	0.2	0.48	***AAA(Lys)***	***1.66***	1.84	1.49
GTA(Val)	***1.52***	1.68	1.35	AAG(Lys)	0.34	0.16	0.51
GTG(Val)	0.56	0.26	0.85	GAT(Asp)	***1.58***	1.46	1.59
TCT(Ser)	***1.41***	1.61	1.23	GAC(Asp)	0.42	0.54	0.41
TCC(Ser)	0.23	0.06	0.32	***GAA(Glu)***	***1.68***	1.81	1.55
TCA(Ser)	***1.87***	2.54	1.67	GAG(Glu)	0.32	0.19	0.45
TCG(Ser)	0.32	0.10	0.58	***TGT(Cys)***	***1.63***	1.76	1.63
AGT(Ser)	***1.66***	1.2	1.69	TGC(Cys)	0.37	0.24	0.37
AGC(Ser)	0.5	0.49	0.52	**CGT(Arg)**	**2.23**	3.74	1.52
CCT(Pro)	**1.47**	1.39	1.44	CGC(Arg)	0.53	0.48	0.56
CCC(Pro)	0.17	0.03	0.26	CGA(Arg)	0.55	0.1	0.97
CCA(Pro)	**1.95**	2.38	1.74	CGG(Arg)	0.09	0	0.22
CCG(Pro)	0.41	0.19	0.56	AGA(Arg)	***2.32***	1.66	2.11
ACT(Thr)	***1.28***	1.63	1.14	AGG(Arg)	0.28	0.03	0.62
ACC(Thr)	0.23	0.04	0.37	**GGT(Gly)**	**2.09**	2.54	1.81
ACA(Thr)	***1.92***	2.05	1.70	**GGC(Gly)**	0.66	0.66	0.64
ACG(Thr)	0.57	0.28	0.79	GGA(Gly)	0.93	0.71	0.99
				GGG(Gly)	0.32	0.09	0.57

Note: Bold text indicates preferentially used codons (total RSCU>1.00). Red highlights indicate codons that are AT rich and preferentially used. There was a statistically significant difference between RSCUPHX and RSCUPLX (P<.01). The italics indicate the optimum codons, which are those that are preferentially used by potentially highly expressed genes relative to potentially weakly expressed ones [see Section 2.4]. Green abbreviations have been used to indicate GC rich optimum codons: RSCU stands for relative synonymous codon use (cumulative); PHX for potentially highly expressed genes; PLX for potentially lowly expressed genes; and Aa for amino acids. [a]codon that each amino acid prefers.

5.9. Pathogens tends to favor amino acid with low biosynthetic costs

In our comprehensive analysis of amino acid usage across different categories of the *Staphylococcus* genus that includes Coagulase-positive staphylococci (CoPS), Coagulase-negative staphylococci (CoNS), and Coagulase-variable species (**Fig 6**). We identified clear patterns that reveal both preferred and avoided amino acids within these groups. Specifically, leucine, isoleucine, and lysine emerged as the most frequently utilized amino acids in CoPS and CoNS. This suggests that these amino acids may play a crucial role in the metabolic activities, protein synthesis, or structural integrity of these bacterial groups. Leucine, in particular, stands out as a commonly favored amino acid across all three categories, indicating its fundamental importance for the genus as a whole. Leucine's role in protein stability and its involvement in energy metabolism might explain its high usage. In Coagulase-variable species, however, we observed a preference for alanine and leucine, suggesting that alanine may play a more significant role in the biology of these variable strains, possibly due to its contribution to smaller, more flexible proteins or its role in cellular energetics.

On the other hand, our analysis revealed a consistent avoidance of certain amino acids. In both CoPS and CoNS, cysteine, methionine, proline, histidine, and tryptophan were the most avoided. Interestingly, for Coagulase-variable species, threonine, methionine, arginine, cysteine, and tryptophan were the least favored. A notable observation is the consistent underrepresentation of sulfur-containing amino acids such as methionine and cysteine across all categories. This avoidance could be linked to the high energetic cost or regulatory complexity associated with the synthesis and incorporation of sulfur-containing amino acids. Additionally, cysteine’s involvement in disulfide bond formation, which is crucial in maintaining protein structure, may be less necessary or less efficient under the environmental conditions these bacteria encounter. Similarly, the avoidance of methionine could be related to its role as a start codon in protein synthesis, where stringent regulation is required, or its involvement in processes like methylation, which might be minimized in certain bacterial environments. Furthermore, amino acids like proline, histidine, and tryptophan were also less frequently used. Proline's unique structure often leads to bends or disruptions in protein folding, which

might explain its avoidance in bacterial proteins that require stable, elongated structures. Histidine, due to its role in enzyme active sites and pH regulation, might be used selectively based on environmental pH levels, while tryptophan, being one of the rarest amino acids in nature, could be avoided due to its biosynthetic cost or limited availability in nutrient-poor environments. Threonine and arginine avoidance in Coagulase-variable species might point to specific metabolic or regulatory differences in these strains, which could reflect their adaptation to different ecological niches or physiological requirements compared to CoPS and CoNS. Overall, the patterns of amino acid usage in the *Staphylococcus* genus provide valuable insights into the evolutionary pressures and metabolic constraints faced by these bacteria. The preference for certain amino acids like leucine across all categories and the avoidance of others such as methionine and cysteine may reflect a combination of factors, including energy efficiency, protein stability, and environmental adaptation. Understanding these patterns can offer deeper insight into the functional and adaptive strategies of *Staphylococcus* species, which could be crucial for developing new antimicrobial strategies or understanding bacterial resilience in various environments.

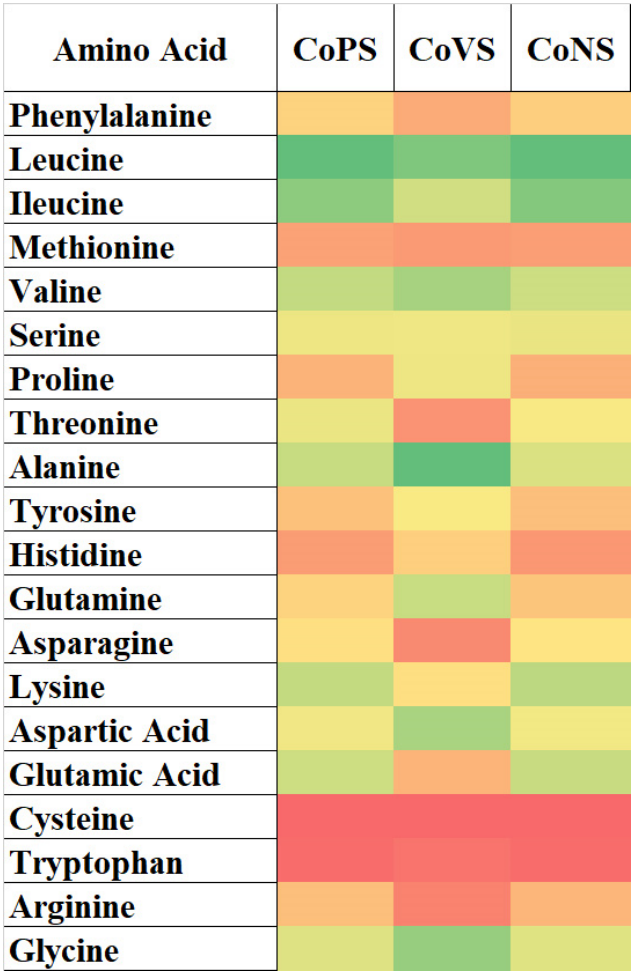


Fig. 6: Heatmap of Amino Acids Usage of Coagulase positive *Staphylococcus* (CoPS), Coagulase Negative *Staphylococcus* (CoNS) and Coagulase variable *Staphylococcus* (CoVS) groups. (green indicates high usage of amino acid and red indicated low usage of amino acid)

The selection of simpler amino acids like lysine, isoleucine, and leucine as the most preferred across the *Staphylococcus* genus likely reflects an adaptive strategy for energy efficiency. These amino acids are less complex and energetically cheaper to synthesize, making them favorable choices for maintaining cellular processes without placing excessive metabolic demands on the organism. This is particularly important in the context of protein secretion, where the secretome—composed of proteins destined to be expelled from the cell—requires substantial energy investment. Since these secreted proteins, once outside the cell, have little to no chance of being recycled, the bacteria must conserve resources by favoring simpler amino acids in their production.

Conversely, bulkier and more metabolically expensive amino acids like tryptophan, methionine, cysteine, histidine, and proline are avoided, likely because their biosynthesis and incorporation into proteins are costlier and more complex. By minimizing the use of these bulky amino acids, *Staphylococcus* species reduce the overall energetic burden on the cell, especially during protein secretion processes, which are crucial for survival and pathogenicity. This economical use of simpler amino acids thus contributes to the bacteria's cellular fitness, promoting more efficient growth and adaptation in various environments. This strategy highlights the balance between functional protein production and metabolic conservation, ensuring that resources are allocated efficiently to support the bacteria's survival and proliferation.

5.10. Hydrophobicity and Aromaticity shape amino acid usage

Our GRAVY and AROMO analysis of amino acid usage in the *Staphylococcus* genus reveals a complex, multifactorial influence shaping the patterns of amino acid preference. The GRAVY index, which measures the hydrophobicity of proteins, showed a clear correlation between the average hydropathicity of amino acids and their relative usage (RAAU). This relationship highlights the importance of hydrophobic residues in protein structures, particularly for proteins involved in membrane interactions or secretion, where stability in hydrophobic or aqueous environments is crucial. Hydrophobic amino acids tend to be positioned in the core of proteins, playing a vital role in maintaining structural integrity, while hydrophilic ones are more often found on the surface, interacting with the surrounding environment.

The analysis also identified a significant correlation between the average aromaticity of amino acids (AROMO) and RAAU. Aromatic amino acids, though less frequently used due to their biosynthetic complexity, play critical roles in stabilizing protein structures through stacking interactions and are essential in specific protein-protein interactions. Their selective incorporation into the proteome underscores the need for functional precision, despite the higher metabolic costs associated with their synthesis.

These findings suggest that both hydropathicity and aromaticity exert significant influences on amino acid usage, reflecting a balance between the need for protein stability, functional efficiency, and metabolic economy. This balance allows

Staphylococcus species to optimize their protein composition in ways that support their adaptability and survival across diverse environments. (**Table 5**).

Table 5: Hydrophobicity and aromaticity shape amino acid usage in *Staphylococcus* species

Organism		Nc	GC3s	GC	L_aa	Gravy	Aromo	CAI
Staphylococcus agnetis	**Axis 1 RAAU**	-.188**	-.284**	-.082**	-.094**	-.895**	-.433**	.412**
	Axis 2 RAAU	.308**	.070**	-.648**	-.239**	-.125**	.533**	-.269**
	Axis 1 RSCU	-.588**	-.302**	.188**	.086**	-.139**	-.280**	.740**
	Axis 2 RSCU	.249**	.453**	.362**	.206**	.147**	-0.020	-.398**
Staphylococcus argenteus	**Axis 1 RAAU**	-.100**	-.121**	-0.033	-.098**	-.910**	-.376**	.292**
	Axis 2 RAAU	.256**	.190**	-.764**	-.310**	-.212**	.501**	-.267**
	Axis 1 RSCU	-.412**	-.208**	.350**	.118**	-.046*	-.237**	.662**
	Axis 2 RSCU	.325**	.188**	.082**	.094**	.093**	.058**	-.374**
Staphylococcus arlettae	**Axis 1 RAAU**	-.098**	-.132**	-.082**	-.164**	-.913**	-.326**	.291**
	Axis 2 RAAU	.312**	.234**	-.698**	-.239**	-.130**	.581**	-.250**
	Axis 1 RSCU	.451**	.254**	-.260**	-.096**	.119**	.246**	-.656**
	Axis 2 RSCU	.320**	.373**	.183**	.064**	.102**	.077**	-.336**
Staphylococcus aureus	**Axis 1 RAAU**	-.098**	-.132**	-.082**	-.164**	-.913**	-.326**	.291**
	Axis 2 RAAU	.312**	.234**	-.698**	-.239**	-.130**	.581**	-.250**
	Axis 1 RSCU	.451**	.254**	-.260**	-.096**	.119**	.246**	-.656**
	Axis 2 RSCU	.320**	.373**	.183**	.064**	.102**	.077**	-.336**
Staphylococcus auricularis	**Axis 1 RAAU**	-.199**	-.419**	-.165**	-.081**	-.900**	-.449**	.346**
	Axis 2 RAAU	.308**	.141**	-.514**	-.251**	-.178**	.467**	-.385**
	Axis 1 RSCU	.674**	.527**	.057*	-0.007	.180**	.262**	-.901**

	Axis 2 RSCU	-.289**	-.530**	-.393**	-.256**	-.102**	-.064**	.297**
Staphylococcus capitis	**Axis 1 RAAU**	.119**	.170**	.097**	.048*	.915**	.321**	-.216**
	Axis 2 RAAU	.221**	.150**	-.791**	-.217**	-.119**	.534**	-.277**
	Axis 1 RSCU	-.356**	-.170**	.342**	.114**	-.112**	-.266**	.500**
	Axis 2 RSCU	-.171**	-.249**	-.240**	-.087**	-0.002	0.027	.184**
Staphylococcus caprae	**Axis 1 RAAU**	-.136**	-.180**	-.077**	-.121**	-.904**	-.396**	.291**
	Axis 2 RAAU	.231**	.073**	-.804**	-.287**	-.178**	.519**	-.268**
	Axis 1 RSCU	-.497**	-.239**	.336**	.127**	-.079**	-.287**	.717**
	Axis 2 RSCU	0.016	.164**	.345**	.120**	-.062**	-.146**	.127**
Staphylococcus carnosus	**Axis 1 RAAU**	.138**	.204**	.128**	.140**	.909**	.359**	-.255**
	Axis 2 RAAU	.315**	.076**	-.717**	-.198**	-.119**	.500**	-.114**
	Axis 1 RSCU	.634**	.285**	-.266**	-.077**	.091**	.271**	-.524**
	Axis 2 RSCU	-.313**	-.482**	-.395**	-.180**	-.043*	-0.018	.442**
Staphylococcus chromogenes	**Axis 1 RAAU**	.145**	.270**	.207**	.153**	.893**	.351**	-.188**
	Axis 2 RAAU	.234**	.044*	-.633**	-.288**	-.144**	.538**	-.365**
	Axis 1 RSCU	.612**	.466**	0.013	-0.011	.135**	.255**	-.668**
	Axis 2 RSCU	-.246**	-.522**	-.497**	-.250**	-.216**	0.022	.172**
Staphylococcus cohnii	**Axis 1 RAAU**	-.072**	-.143**	-.052*	-.134**	-.902**	-.364**	.239**
	Axis 2 RAAU	-.276**	-.157**	.769**	.230**	.103**	-.564**	.260**
	Axis 1 RSCU	.390**	.228**	-.293**	-0.040	.104**	.238**	-.620**
	Axis 2 RSCU	.421**	.448**	.180**	-0.035	-0.023	.042*	-.337**
Staphylococcus condimenti	**Axis 1 RAAU**	-.131**	-.256**	-.180**	-.088**	-.890**	-.337**	.270**
	Axis 2 RAAU	-.288**	-.043*	.694**	.230**	.225**	-.476**	.262**

	Axis 1 RSCU	-.647**	-.690**	-.153**	.085**	-.122**	-.183**	.799**
	Axis 2 RSCU	-.338**	.080**	.426**	.105**	-0.016	-.218**	.470**
Staphylococcus delphini	**Axis 1 RAAU**	.200**	.377**	.201**	.129**	.877**	.392**	-.360**
	Axis 2 RAAU	.069**	0.016	-.088**	-.163**	-.296**	.177**	-.075**
	Axis 1 RSCU	-.618**	-.543**	-.106**	0.014	-.133**	-.219**	.883**
	Axis 2 RSCU	-.149**	.364**	.473**	.323**	.130**	-.103**	-.095**
Staphylococcus devriesei	**Axis 1 RAAU**	.143**	.210**	.135**	.162**	.896**	.352**	-.272**
	Axis 2 RAAU	-.307**	-.107**	.756**	.245**	.141**	-.520**	.334**
	Axis 1 RSCU	.541**	.245**	-.346**	-.135**	.088**	.281**	-.727**
	Axis 2 RSCU	.439**	.501**	.335**	.190**	.101**	.047*	-.438**
Staphylococcus edaphicus	**Axis 1 RAAU**	-.116**	-.230**	-.083**	-.134**	-.909**	-.387**	.298**
	Axis 2 RAAU	-.133**	-.073**	.506**	.085**	-.115**	-.210**	.130**
	Axis 1 RSCU	-.294**	-.137**	.293**	.066**	-.109**	-.215**	.562**
	Axis 2 RSCU	.414**	.409**	.193**	.117**	.055**	.043*	-.373**
Staphylococcus epidermidis	**Axis 1 RAAU**	-.137**	-.153**	-0.017	-.046*	-.906**	-.409**	.300**
	Axis 2 RAAU	.196**	.199**	-.813**	-.303**	-.106**	.538**	-.256**
	Axis 1 RSCU	-.310**	-.184**	.317**	.070**	-.087**	-.234**	.586**
	Axis 2 RSCU	-.098**	-.141**	-.297**	-.117**	0.022	.097**	-.046*
Staphylococcus equorum	**Axis 1 RAAU**	.051**	.061**	.194**	.216**	.910**	.222**	-.170**
	Axis 2 RAAU	.257**	.180**	-.795**	-.260**	-0.027	.590**	-.314**
	Axis 1 RSCU	-.346**	-.235**	.330**	.181**	-0.018	-.246**	.629**
	Axis 2 RSCU	.371**	.298**	.162**	0.031	.055**	0.014	-.338**
Staphylococcus felis	**Axis 1 RAAU**	-.156**	-.170**	-0.023	-.092**	-.882**	-.400**	.326**

	Axis 2 RAAU	-.198**	-.047*	.741**	.193**	.171**	-.518**	.194**
	Axis 1 RSCU	.316**	0.040	-.373**	0.012	.088**	.224**	-.515**
	Axis 2 RSCU	.448**	.521**	.306**	.098**	.155**	.094**	-.582**
Staphylococcus fleurettii	**Axis 1 RAAU**	0.009	-.066**	-.209**	-.156**	-.894**	-.260**	.137**
	Axis 2 RAAU	-.280**	-.218**	.800**	.111**	0.026	-.592**	.376**
	Axis 1 RSCU	-.246**	-.120**	.409**	0.036	-.082**	-.228**	.586**
	Axis 2 RSCU	.510**	.453**	.117**	-.159**	-.047*	.060**	-.404**
Staphylococcus gallinarum	**Axis 1 RAAU**	-.149**	-.137**	0.025	-.153**	-.904**	-.410**	.329**
	Axis 2 RAAU	-.269**	-.118**	.762**	.251**	.143**	-.533**	.244**
	Axis 1 RSCU	-.380**	-.219**	.279**	.092**	-.107**	-.230**	.685**
	Axis 2 RSCU	-.341**	-.334**	-.120**	0.010	-.045*	-.047*	.167**
Staphylococcus haemolyticus	**Axis 1 RAAU**	.072**	.175**	.194**	.146**	.893**	.319**	-.219**
	Axis 2 RAAU	-.354**	-.278**	.663**	.204**	.162**	-.518**	.374**
	Axis 1 RSCU	.612**	.428**	-.208**	-.128**	.039*	.208**	-.797**
	Axis 2 RSCU	-.268**	-.379**	-.288**	-.094**	-0.034	-0.015	.269**
Staphylococcus hominis	**Axis 1 RAAU**	-.081**	-.115**	-.133**	-.106**	-.899**	-.348**	.252**
	Axis 2 RAAU	-.259**	-.160**	.819**	.210**	.108**	-.512**	.307**
	Axis 1 RSCU	-.340**	-.220**	.385**	.177**	-.053**	-.235**	.598**
	Axis 2 RSCU	.232**	.055**	-.197**	-0.017	.099**	.096**	-.339**
Staphylococcus hyicus	**Axis 1 RAAU**	-.165**	-.255**	-.093**	-.115**	-.895**	-.422**	.367**
	Axis 2 RAAU	.288**	.085**	-.606**	-.259**	-.165**	.533**	-.300**
	Axis 1 RSCU	.511**	.202**	-.276**	-.094**	.072**	.262**	-.689**
	Axis 2 RSCU	-.302**	-.429**	-.339**	-.169**	-.168**	-0.038	.466**

Staphylococcus intermedius	**Axis 1 RAAU**	-.162**	-.356**	-.264**	-.179**	-.891**	-.359**	.328**
	Axis 2 RAAU	-.058**	0.006	.167**	.206**	.322**	-.203**	.038*
	Axis 1 RSCU	.222**	-.285**	-.544**	-.355**	-.174**	.177**	-.185**
	Axis 2 RSCU	.580**	.553**	.229**	.067**	.136**	.154**	-.813**
Staphylococcus kloosii	**Axis 1 RAAU**	.151**	.190**	-.044*	.127**	.896**	.390**	-.357**
	Axis 2 RAAU	-.314**	-.142**	.796**	.227**	.180**	-.529**	.280**
	Axis 1 RSCU	.518**	.275**	-.323**	-.086**	.132**	.272**	-.767**
	Axis 2 RSCU	.338**	.351**	.067**	0.004	0.006	.073**	-.240**
Staphylococcus lentus	**Axis 1 RAAU**	.054**	.129**	.170**	.174**	.899**	.323**	-.217**
	Axis 2 RAAU	-.280**	-.165**	.828**	.192**	.055**	-.561**	.377**
	Axis 1 RSCU	-.238**	-.107**	.386**	.063**	-.129**	-.207**	.640**
	Axis 2 RSCU	.416**	.325**	.077**	-.089**	-0.016	.049*	-.306**
Staphylococcus lugdunensis	**Axis 1 RAAU**	.161**	.237**	0.025	.098**	.894**	.428**	-.307**
	Axis 2 RAAU	.267**	.154**	-.737**	-.248**	-.181**	.506**	-.210**
	Axis 1 RSCU	.422**	.233**	-.284**	-.104**	.076**	.239**	-.570**
	Axis 2 RSCU	-.413**	-.503**	-.271**	-.169**	-.076**	-.051**	.488**
Staphylococcus lutrae	**Axis 1 RAAU**	-.257**	-.321**	-0.033	-.079**	-.803**	-.501**	.374**
	Axis 2 RAAU	-.145**	.106**	.667**	.292**	.389**	-.385**	.071**
	Axis 1 RSCU	-.067**	-.430**	-.452**	-.218**	-.105**	.089**	.188**
	Axis 2 RSCU	.519**	.463**	.109**	.069**	.138**	.161**	-.777**
Staphylococcus massiliensis	**Axis 1 RAAU**	.261**	.261**	.046*	.073**	.886**	.414**	-.364**
	Axis 2 RAAU	-.268**	-.063**	.601**	.283**	.214**	-.501**	.351**
	Axis 1 RSCU	-.689**	-.507**	-.042*	.063**	-.163**	-.288**	.916**

	Axis 2 RSCU	-.165**	-.583**	-.619**	-.245**	-0.037	.071**	.135**
Staphylococcus microti	**Axis 1 RAAU**	.121**	.250**	.048*	-0.015	.671**	.444**	-.300**
	Axis 2 RAAU	.111**	.176**	-0.016	.113**	.758**	.371**	-.147**
	Axis 1 RSCU	.550**	.050*	-.345**	-.210**	-.091**	.245**	-.601**
	Axis 2 RSCU	.276**	.568**	.407**	.179**	.175**	.046*	-.628**
Staphylococcus muscae	**Axis 1 RAAU**	.191**	.231**	0.028	.094**	.839**	.466**	-.293**
	Axis 2 RAAU	-.350**	-.075**	.573**	.182**	.235**	-.418**	.309**
	Axis 1 RSCU	.597**	.268**	-.126**	-.053*	.076**	.216**	-.821**
	Axis 2 RSCU	-.049*	-.429**	-.394**	-.215**	-.121**	.047*	.276**
Staphylococcus nepalensis	**Axis 1 RAAU**	-.113**	-.168**	-.040*	-.154**	-.903**	-.393**	.320**
	Axis 2 RAAU	.253**	0.016	-.745**	-.119**	-.121**	.541**	-.252**
	Axis 1 RSCU	-.319**	.145**	.465**	-.146**	-.110**	-.208**	.499**
	Axis 2 RSCU	.392**	.312**	.071**	.069**	.048**	.072**	-.320**
Staphylococcus pasteuri	**Axis 1 RAAU**	-.067**	-.145**	-.076**	-.104**	-.899**	-.391**	.252**
	Axis 2 RAAU	-.235**	-.056**	.822**	.064**	.097**	-.508**	.170**
	Axis 1 RSCU	-.681**	-.730**	-.114**	.191**	-.063**	-.130**	.831**
	Axis 2 RSCU	.246**	-.049*	-.422**	-0.009	.110**	.222**	-.423**
Staphylococcus petrasii subsp. croceilyticus	**Axis 1 RAAU**	-.138*	-0.127	0.079	0.051	-.473**	-.424**	.241**
	Axis 2 RAAU	-0.005	-.234**	-.384**	-0.068	-.769**	0.101	.237**
	Axis 1 RSCU	.462**	.246**	-.223**	-.176**	-0.029	.173**	-.593**
	Axis 2 RSCU	.511**	.290**	-.136*	.172**	0.039	.308**	-.624**
Staphylococcus pettenkoferi	**Axis 1 RAAU**	.243**	.315**	.188**	.125**	.901**	.407**	-.379**
	Axis 2 RAAU	-.173**	.061**	.545**	.300**	.250**	-.457**	.239**

	Axis 1 RSCU	.651**	.584**	.231**	0.020	.181**	.251**	-.912**
	Axis 2 RSCU	.041*	-.531**	-.604**	-.313**	-.085**	.132**	0.039
Staphylococcus piscifermentans	**Axis 1 RAAU**	-.103**	-.186**	-.089**	-.125**	-.912**	-.338**	.286**
	Axis 2 RAAU	.272**	.141**	-.795**	-.233**	-.089**	.580**	-.302**
	Axis 1 RSCU	-.401**	-.191**	.340**	.080**	-.100**	-.245**	.656**
	Axis 2 RSCU	-.334**	-.282**	-.113**	0.029	-0.014	-0.022	.267**
Staphylococcus pseudintermedius	**Axis 1 RAAU**	.169**	.294**	.100**	.112**	.878**	.412**	-.335**
	Axis 2 RAAU	-.243**	0.032	.749**	.269**	.216**	-.544**	.208**
	Axis 1 RSCU	.523**	.151**	-.286**	-.166**	.043*	.282**	-.643**
	Axis 2 RSCU	.354**	.643**	.429**	.186**	.149**	0.020	-.618**
Staphylococcus pseudoxylosus	**Axis 1 RAAU**	.043*	.122**	.086**	.181**	.914**	.318**	-.278**
	Axis 2 RAAU	.301**	.208**	-.746**	-.266**	-.111**	.593**	-.297**
	Axis 1 RSCU	-.360**	-.174**	.315**	.066**	-.102**	-.210**	.670**
	Axis 2 RSCU	-.299**	-.236**	-.170**	-0.035	-0.017	-0.028	.197**
Staphylococcus rostri	**Axis 1 RAAU**	-.184**	-.237**	-.058**	-.125**	-.859**	-.456**	.315**
	Axis 2 RAAU	.170**	.148**	.057**	-.153**	-.260**	.145**	-.213**
	Axis 1 RSCU	-.579**	-.266**	.133**	.131**	-0.007	-.243**	.806**
	Axis 2 RSCU	.169**	.625**	.519**	.272**	.160**	-0.007	-.433**
Staphylococcus saccharolyticus	**Axis 1 RAAU**	.077**	.084**	.053*	.120**	.906**	.349**	-.274**
	Axis 2 RAAU	-.331**	-.206**	.662**	0.038	0.007	-.357**	.378**
	Axis 1 RSCU	-.470**	-.225**	.399**	.099**	-.121**	-.270**	.775**
	Axis 2 RSCU	-.194**	-.253**	-.177**	0.004	.073**	0.025	.080**
Staphylococcus saprophyticus	**Axis 1 RAAU**	-.061**	-.130**	-.057**	-.088**	-.902**	-.359**	.241**

	Axis 2 RAAU	.272**	.154**	-.782**	-.283**	-.162**	.503**	-.261**
	Axis 1 RSCU	.343**	.203**	-.323**	-.072**	.077**	.216**	-.587**
	Axis 2 RSCU	0.019	-.054**	-.103**	-.049*	.052**	0.001	-0.005
Staphylococcus schleiferi	**Axis 1 RAAU**	-.151**	-.201**	-.107**	-.124**	-.895**	-.387**	.295**
	Axis 2 RAAU	.236**	.090**	-.672**	-.276**	-.199**	.517**	-.237**
	Axis 1 RSCU	-.516**	-.339**	.134**	.092**	-.089**	-.238**	.744**
	Axis 2 RSCU	-.449**	-.585**	-.493**	-.127**	-.121**	-0.021	.464**
Staphylococcus sciuri	**Axis 1 RAAU**	.036*	.128**	.210**	.147**	.895**	.309**	-.202**
	Axis 2 RAAU	-.329**	-.237**	.755**	.177**	.111**	-.551**	.428**
	Axis 1 RSCU	-.363**	-.194**	.377**	.062**	-.055**	-.235**	.689**
	Axis 2 RSCU	.425**	.380**	.146**	-.163**	-.038*	.048**	-.345**
Staphylococcus simiae	**Axis 1 RAAU**	.124**	.179**	0.016	.114**	.882**	.418**	-.278**
	Axis 2 RAAU	-.281**	-.222**	.730**	.295**	.212**	-.490**	.290**
	Axis 1 RSCU	.464**	.322**	-.267**	-.146**	.052*	.192**	-.712**
	Axis 2 RSCU	-.080**	-.087**	-.156**	-.081**	-0.004	.046*	-.041*
Staphylococcus simulans	**Axis 1 RAAU**	.291**	.379**	-0.016	0.038	.821**	.536**	-.432**
	Axis 2 RAAU	.252**	-.058**	-.573**	-.279**	-.431**	.236**	-.238**
	Axis 1 RSCU	-.630**	-.184**	.314**	.155**	-.130**	-.271**	.849**
	Axis 2 RSCU	.227**	.451**	.279**	.170**	.105**	.074**	-.326**
Staphylococcus succinus	**Axis 1 RAAU**	-.127**	-.244**	-.194**	-.058**	-.893**	-.296**	.338**
	Axis 2 RAAU	-.082**	.242**	.883**	-.081**	.090**	-.532**	-.205**
	Axis 1 RSCU	-.449**	-.809**	-.545**	.282**	-.099**	.062**	.632**
	Axis 2 RSCU	-.359**	-.126**	.188**	.055**	-.126**	-.167**	.492**

Staphylococcus vitulinus	**Axis 1 RAAU**	0.010	.144**	.188**	.134**	.892**	.323**	-.229**
	Axis 2 RAAU	.314**	.278**	-.775**	-.218**	-0.025	.590**	-.374**
	Axis 1 RSCU	-.379**	-.279**	.335**	.114**	-.135**	-.254**	.689**
	Axis 2 RSCU	.320**	.293**	.224**	-.058**	-.039*	-0.021	-.134**
Staphylococcus warneri	**Axis 1 RAAU**	.095**	.146**	.080**	.138**	.908**	.379**	-.280**
	Axis 2 RAAU	.362**	.217**	-.700**	-.267**	-.174**	.477**	-.388**
	Axis 1 RSCU	.524**	.321**	-.285**	-.106**	.119**	.250**	-.784**
	Axis 2 RSCU	.327**	.353**	.177**	0.010	-0.023	0.035	-.240**
Staphylococcus xylosus	**Axis 1 RAAU**	.089**	.178**	.198**	.152**	.917**	.301**	-.276**
	Axis 2 RAAU	-.208**	-.112**	.837**	.275**	.038*	-.593**	.270**
	Axis 1 RSCU	-.308**	-.115**	.363**	.103**	-.108**	-.230**	.622**
	Axis 2 RSCU	.412**	.365**	.196**	.043*	.041*	0.009	-.286**

5.11. Corresponding Analysis differentiates between CoPS, CoNS and Co-variables

Coagulase-Positive Staphylococci (CoPS), Coagulase-Negative Staphylococci (CoNS), and other variable species are three different groupings within the Staphylococcus genus that exhibit notable genomic variations, according to the analysis of data plots comparing GC12 vs. GC3s, ENc vs. GC3, and ENc vs. GC12 (**Fig. 7.**). These differences are reflected in the distinct clustering patterns observed across all three plots, indicating that each group possesses unique genomic signatures. Important markers of these genomic fingerprints include the effective number of codons (ENc), the GC content at the first and second codon locations (GC12), and the GC content at the third codon position (GC3s). In particular, ENc indicates the level of codon usage variety, whereas GC3s variability is frequently associated with codon usage bias, which might disclose evolutionary forces forming the genome. The clustering of CoPS, CoNS, and variable species in these plots suggests that the genomic architecture of each group

has evolved under different selective pressures, likely reflecting their distinct biological roles, ecological niches, and pathogenic potential.

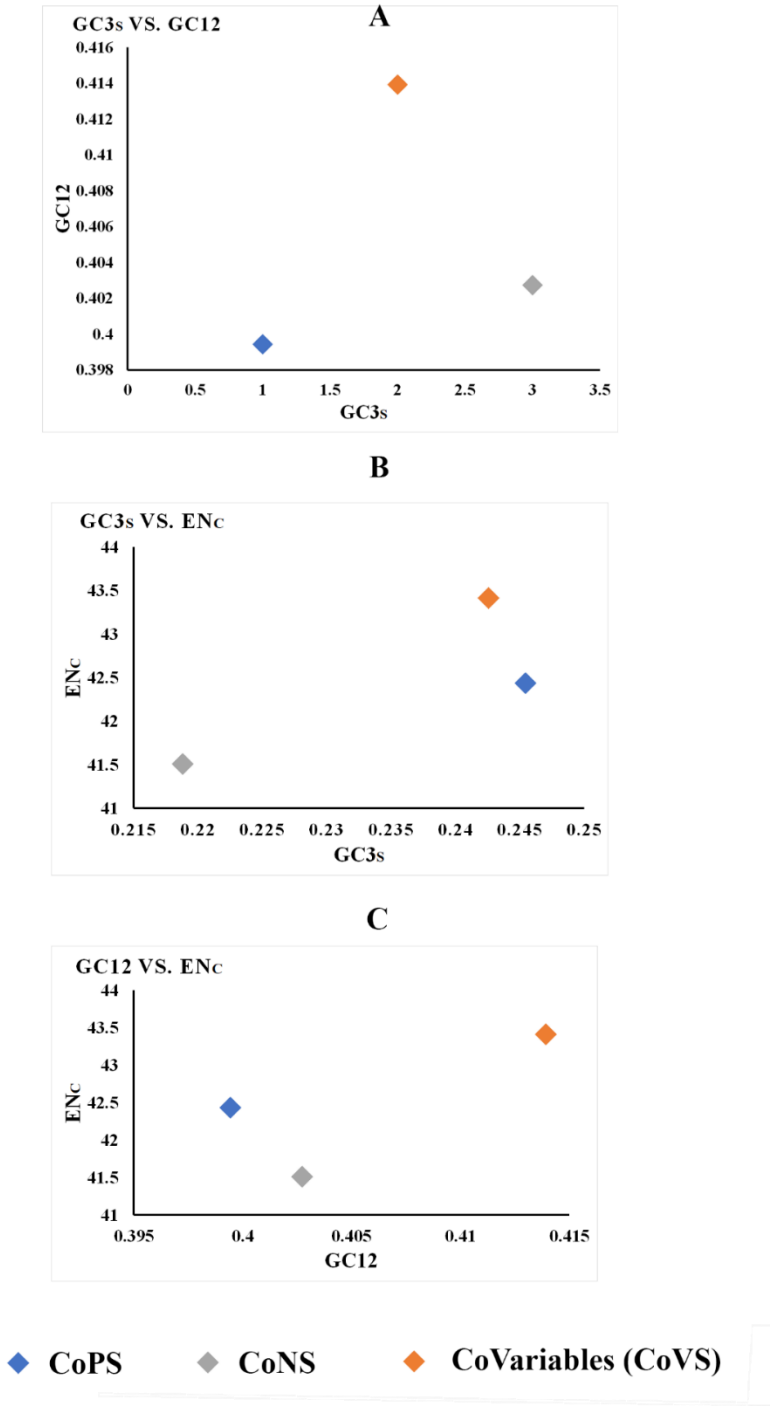


Fig. 7: Cumulative mean of CoPS, CoNS and CoVS A). GC12 vs GC3 B). ENc vs GC3s C). ENc vs GC12

These findings are further reinforced by evolutionary studies, where similar clustering patterns were observed in phylogenetic analyses, showing that CoPS, CoNS, and variable species tend to form separate clades. This indicates a deeper evolutionary divergence among these groups, particularly in relation to their pathogenic traits. Pathogenic species, often associated with CoPS, form distinct clusters, reflecting shifts in both genomic composition and codon usage bias that likely support their ability to cause disease. For instance, pathogenic species may exhibit higher codon usage bias, which could optimize the expression of virulence factors under host conditions. The distinct clustering of CoPS and CoNS reflects not only their genomic differences but also their divergent evolutionary paths, with CoPS typically associated with more virulent and pathogenic traits compared to CoNS, which are often commensal but can also be opportunistic pathogens. The clustering of variable species further emphasizes the complexity of genomic adaptation within the *Staphylococcus* genus, where compositional properties such as GC content and codon usage can serve as markers of evolutionary and functional divergence. This classification system, grounded in both compositional and evolutionary data, offers new insights into how genomic and codon usage shifts have driven the evolution of pathogenic characteristics within the genus. It also provides a framework for further studies on the molecular mechanisms underlying these shifts and their implications for virulence, adaptation, and antibiotic resistance.

5.12 Validation of codon usage bias in selected species by genome sequencing

To verify the CUB findings, we integrated *in vivo* genome sequencing with comprehensive *in silico* bioinformatics analyses for *Staphylococcus* hominis (GenBank: GCA_004329095) and *Staphylococcus* intermedius (GenBank: GCA_900458545). The genomic data quality was rigorously evaluated using the CheckM2 tool to confirm both genome completeness and absence of contamination, critical for accurate downstream analysis. Our findings demonstrated a high degree of completeness for both species, with S. hominis and S. intermedius achieving near-perfect genome coverage at 99.99%. Additionally, contamination levels were notably low, measured at 0.04% for S. hominis and 0.25% for S. intermedius, underscoring the reliability of the sequencing process. The consistency of these results with expected genome assembly metrics, such as contig size and total genome length, further validates

the robustness of the genomic data and supports its use in comparative analyses. These measures assure that the genomes are well-assembled and meet quality thresholds necessary for reliable bioinformatics and functional annotation studies.

A key factor in codon usage bias, GC content, was analyzed across both species to evaluate the alignment between *in vivo* and in silico data. In S. hominis GC content derived from *in vivo* sequencing was measured at 31.39%, a value closely corroborated by *in silico* analysis, which reported a similar GC content of 31.4%. Likewise, for S. intermedius, *in vivo* sequencing revealed a GC content of 37.62%, with a marginally higher GC content of 38.03% observed through *in silico* methods. This near-identical GC content across methodologies indicates the accuracy and robustness of the bioinformatics approach in reflecting *in vivo* genomic characteristics. This alignment supports the validity of *in silico* GC content analysis as a proxy for *in vivo* findings, reinforcing the reliability of computational analyses in capturing essential genomic properties that affect codon preference, gene expression, and evolutionary adaptation across related bacterial species (**Fig. 8.**).

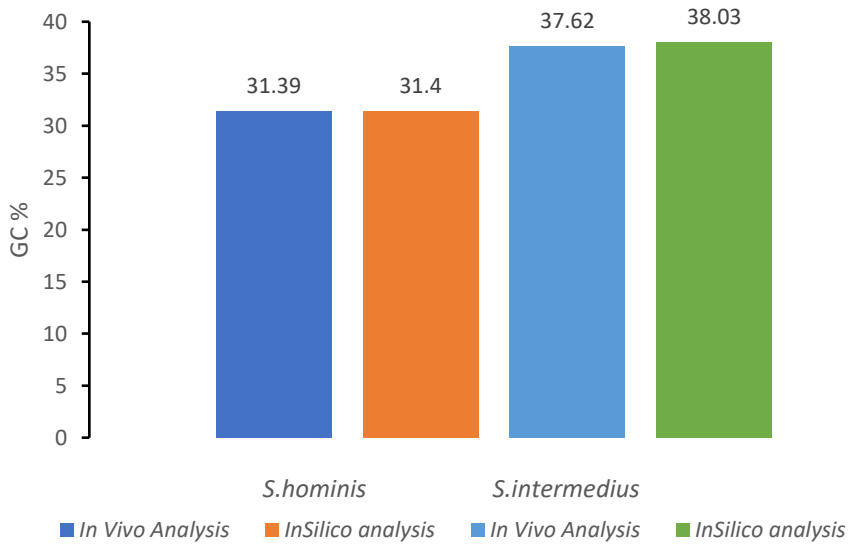


Fig. 8. GC% Comparison of *in silico* and *in vivo* analysis

5.13. Discussion

The *Staphylococcus* genus, composed of Gram-positive bacteria, holds significant importance in microbiology and medicine for its involvement in human infections and resistance to antibiotics. Notably, pathogens like *Staphylococcus aureus* and *Staphylococcus epidermidis* are known to cause diverse diseases, ranging from superficial skin infections to severe systemic illnesses. Understanding the genetic composition and codon usage patterns of *Staphylococcus* species is crucial for elucidating their evolutionary dynamics, adaptive strategies, and pathogenic potential.

In this study, we employed a comprehensive array of analytical methods, including ENc, relative synonymous codon use RSCU, and CAI, to examine codon usage across 48 species within the *Staphylococcus* genus. Additionally, we leveraged neutrality plots, correspondence analysis (COA), and parity plots to gain a thorough understanding of codon usage patterns and their implications for microbial evolution and adaptation. Initially, we observed G-C content of the genus and found variations in genomic G-C content across the 48 species within the *Staphylococcus* genus providing valuable insights into the genetic structure and evolutionary dynamics of these microorganisms. Our findings revealed a range of G-C content from 34.27% to 36% among the examined species. Notably, coagulase-positive species exhibited a higher G-C content of 36%, while coagulase-negative species and covariables displayed lower G-C content values of 34.47% and 34.29%, respectively. Importantly, the differences in G-C content had implications for codon usage preferences across the genus. Additionally, we observed a preference for A/T nucleotides at the third position of codons, with CoPS exhibiting a 72% preference for A/T compared to 73.07% for CoNS and 73.91% for covariables. The prevalence of A/T-rich codons and the bias towards an AT-rich genomic structure in pathogenic *Staphylococcus* species have significant implications for their survival and adaptation. A genome enriched in AT content provides advantages such as increased availability of A/T-based metabolites, which can support growth and sustenance post-infection within the host organism. Furthermore, the presence of less stable AT-rich regions in the genome serves as sites for DNA unwinding and replication initiation, facilitating the replication process. Our analysis of GC3s vs ENc highlights significant codon usage bias across CoPS, CoNS, and co-

variables within the *Staphylococcus* genus. The trend of bias diminishes in the order of co-variables > CoPS > CoNS. Elevated ENc values observed across all three classes of *Staphylococcus* genomes suggest a preference for liberal codons and reduced bias in synonymous codon usage. Plotting GC3s versus ENc in the graph highlights how translational selection shapes patterns of codon utilization. Notably, the decreased gene count seen in CoPS as opposed to CoNS and Coagulase-variables suggests a lesser tRNA load on the pathogen's genomic structure due to a predilection for certain codons. These results provide insight into the adaptive mechanisms used by *Staphylococcus* species to maximize gene expression within their genomic landscape and clarify the intricate interaction between natural selection and mutational pressure that drives codon usage bias.

The neutrality plot between GC3s and GC12 provided insights into the influence of mutational pressure on overall codon usage within the genome. Negative regression coefficient values close to 0 for both CoPS (-0.015) and CoPS (-0.026) suggest a predominant role of natural or translational selection over mutational constraints. Similarly, the positive coefficient value for Coagulase-variables (0.05) indicates the prevalence of natural or translational selection over mutational constraints in this group as well. The analysis of genomic data plots reveals distinct clustering among CoPS, CoNS, and other *Staphylococcus* variables, aligning with evolutionary studies, indicating a shift towards pathogenic traits in genomic composition. These findings highlighted the significant impact of natural or translational selection in constructing codon usage patterns across CoPS, CoNS, and Coagulase variables within the genome. Our analysis of purine and pyrimidine usage bias within intrastent compositions, using parity plot analysis (PR2), offered further insights into the genomic dynamics of *Staphylococcus* species. By plotting the graph between G3s/G3s + C3s (GC bias) and A3s/A3 +T3s (AT bias), we examined the relative distribution of purines and pyrimidines in DNA. Values above 0.5 indicated a preference for purines over pyrimidines, while values below 0.5 suggested a preference for pyrimidines over purines. Our study observed mean AT and GC bias values above 0.5 across the *Staphylococcus* genus, indicating a preference for A over T and G over C at the third codon position. These findings shed light on the underlying mechanisms influencing

nucleotide composition and codon usage patterns in whole genomes. Overexpression of dinucleotide pairs such as GpC, ApC, GpA, and TpG was evident in CoPS, while GpG, CpC, and CpT were least expressed. Similarly, CoNS exhibited overexpression of GpC, ApC, and GpA, with CpC, GpG, and TpC being least abundant. Additionally, CpG was found to be the least abundant dinucleotide pair across both CoPS and CoNS. The underrepresentation of CpG is noteworthy, given its crucial role in initiating various immune responses. Understanding dinucleotide pair expression patterns can provide insights into the pathogenic potential and immune evasion strategies of *Staphylococcus* species. Codon usage bias, crucial for gene expression, was evident across CoPS, CoNS, and co-variables within genus. Notably, TTA (Leu) emerged as the favored codon genus-wide, with CGT (Arg), GGT (Gly), and AGA (Arg) preferred in CoPS. Conversely, CCC (Pro), TCC (Ser), and CTG (Leu) were least preferred for CoPS. This analysis highlights the prevalence of AT-rich codons, particularly TTA, underscoring evolutionary adaptation towards host-dependent energy and nutrition sufficiency, favoring T-rich codons for survivability. Moreover, the exclusive use of TTA for leucine across the genus underscores its significance in *Staphylococcus* genome evolution.

Our analysis of codon pairs and dinucleotide pair expression patterns across the genus unveiled intriguing insights into genomic dynamics. NNG-CNN and NNA-CNN emerged as overrepresented codon pairs, suggesting a prevalence of GpC and ApC dinucleotide pairs at codon pair junctions. Conversely, NNT-CNN and NNG-GNN were underrepresented, indicating a reduced presence of TpC and GpG at these junctions. This pattern reflects the expression trends of dinucleotide pairs across genomes, with GpC and ApC being the most expressed and TpC and GpG the least represented.

Furthermore, substantial relationships were found between CAI and Axis 1 and Axis 2 of relative synonymous codon use (RSCU), as well as between CAI and chromosomal G-C concentration. Greater bias in synonymous codon use was indicated by the larger CAI of CoNS, which had a lower G-C concentration. These results highlight the intricate connection between *Staphylococcus* gene expression methods, chromosomal makeup, and codon use patterns, shedding light on the adaptive mechanisms driving microbial evolution within this genus. The analysis of amino acid usage across

Staphylococcus categories highlighted leucine, isoleucine, and lysine as the most preferred, while cysteine and methionine were avoided. These trends suggest a preference for energetically cheaper amino acids, promoting energy efficiency. Significantly, GRAVY and AROMO analyses revealed correlations between amino acid hydropathicity and relative usage, emphasizing multifactorial influences. Additionally, genomic GC content significantly impacted amino acid preferences across the genus. These findings illuminate adaptive strategies in *Staphylococcus* genomes, favoring efficient amino acid usage to enhance cellular fitness and growth.

CHAPTER 6

6. Conclusion

This study delves into the intricate codon usage patterns and genomic adaptations of 48 species within the *Staphylococcus* genus, shedding light on their evolutionary strategies and implications for pathogenicity and survival. Our analyses encompassed various bioinformatics tools, such as the ENc, RSCU, and CAI, to evaluate the codon usage bias and its contributing factors across different species. The examination began with a comparative analysis of G-C content, revealing notable variation among the species. The G-C content ranged from 34.27% to 36%, illustrating distinct genomic compositions between CoPS and CoNS. CoPS were found to have a higher G-C content, often associated with increased pathogenic traits, while CoNS and coagulase-variable species exhibited slightly lower G-C values. These variations in G-C content are not just numerical but hold significant evolutionary importance, influencing codon preference, gene expression efficiency, and the adaptability of these bacteria in diverse environments.

A key observation from this study is the pronounced bias toward A/T nucleotides at the third position of codons across the genus. CoPS showed a 72% preference for A/T, while CoNS and coagulase-variable species demonstrated even higher A/T preferences at 73.07% and 73.91%, respectively. This A/T enrichment in the genomes of *Staphylococcus* species suggests an adaptive advantage, as an A/T-rich genomic structure provides a more efficient means for DNA unwinding and replication. This adaptation could be particularly beneficial in host environments, where rapid replication and metabolic flexibility are crucial for survival and colonization. The presence of A/T-rich regions could facilitate these processes, making the organism more adept at managing the stressors encountered within a host organism. Moreover, an A/T-rich genome may be linked to the utilization of A/T-based metabolites, contributing to energy efficiency and overall cellular fitness during infection.

Our investigation into the codon usage bias, specifically through the GC3s versus ENc plot, highlighted distinct differences across CoPS, CoNS, and coagulase-variable species. The trend of codon usage bias was most pronounced in coagulase-variable species, followed by CoPS and CoNS. High ENc values across these categories suggest

a tendency toward using a broader range of codons and a reduction in synonymous codon bias. This reduced bias indicates that natural selection may play a more significant role in shaping codon usage than mutational pressure, as evidenced by our neutrality plot analysis between GC3s and GC12. For CoPS and CoNS, negative regression coefficients close to zero (e.g., -0.015 for CoPS and -0.026 for CoNS) imply that translational selection rather than mutational constraints drives codon usage. This pattern underscores the role of translational selection in optimizing gene expression, ensuring that codon usage aligns with tRNA availability and cellular needs.

Further exploration using parity plot analysis (PR2) provided insights into the genomic dynamics related to purine and pyrimidine usage within the *Staphylococcus* genus. By plotting G3s/(G3s + C3s) against A3s/(A3s + T3s), we could evaluate the balance between purines and pyrimidines at the third codon position. Our findings indicated a consistent preference for purines (A and G) over pyrimidines (T and C) across the genus, as reflected by mean bias values above 0.5. This preference for purines suggests an evolutionary strategy that may enhance replication fidelity and transcription efficiency, further supporting the organism's adaptability in various environments. The overexpression of certain dinucleotide pairs, such as GpC, ApC, GpA, and TpG, particularly in CoPS, and their underrepresentation in pairs like GpG and CpC, underscores potential regulatory mechanisms that influence genome stability and gene expression. The notable underrepresentation of CpG, a dinucleotide implicated in immune response activation, suggests an adaptive strategy that may aid immune evasion in pathogenic strains.

To validate these findings, we integrated *in vivo* genome sequencing with comprehensive *in silico* bioinformatics analyses for *Staphylococcus* hominis (GenBank: GCA_004329095) and *Staphylococcus* intermedius (GenBank: GCA_900458545). The genomic data quality was rigorously evaluated using the CheckM2 tool to confirm both genome completeness and absence of contamination, critical for accurate downstream analysis. Our findings demonstrated a high degree of completeness for both species, with S. hominis and S. intermedius achieving near-perfect genome coverage at 99.99%. Additionally, contamination levels were notably low, measured at 0.04% for S. hominis and 0.25% for S. intermedius, underscoring the

reliability of the sequencing process. The consistency of these results with expected genome assembly metrics, such as contig size and total genome length, further validates the robustness of the genomic data and supports its use in comparative analyses. These measures assure that the genomes are well-assembled and meet quality thresholds necessary for reliable bioinformatics and functional annotation studies.

A key factor in codon usage bias, GC content, was analyzed across both species to evaluate the alignment between *in vivo* and *in silico* data. In *S. hominis*, GC content derived from *in vivo* sequencing was measured at 31.39%, a value closely corroborated by *in silico* analysis, which reported a similar GC content of 31.4%. Likewise, *for S. intermedius*, *in vivo* sequencing revealed a GC content of 37.62%, with a marginally higher GC content of 38.03% observed through *in silico* methods. This near-identical GC content across methodologies indicates the accuracy and robustness of the bioinformatics approach in reflecting *in vivo* genomic characteristics. This alignment supports the validity of *in silico* GC content analysis as a proxy for *in vivo* findings, reinforcing the reliability of computational analyses in capturing essential genomic properties that affect codon preference, gene expression, and evolutionary adaptation across related bacterial species.

Codon preference analysis revealed intriguing trends, with TTA (coding for leucine) emerging as the most frequently used codon across the *Staphylococcus* genus. In CoPS, codons such as CGT (Arg), GGT (Gly), and AGA (Arg) were also preferred, while codons like CCC (Pro), TCC (Ser), and CTG (Leu) were least favored. The widespread use of TTA points to an evolutionary adaptation that supports efficient energy utilization, potentially linked to the organism's reliance on host-derived nutrients. The preference for T-rich codons reflects a strategic adaptation that may contribute to efficient translation and protein synthesis, aligning with the genomic AT-rich composition observed in many *Staphylococcus* species.

Our study also examined codon pair and dinucleotide pair expression patterns, revealing significant genomic trends. Overrepresented codon pairs, such as NNG-CNN and NNA-CNN, indicated a preference for GpC and ApC at junctions, while underrepresented pairs, such as NNT-CNN and NNG-GNN, showed reduced

occurrences of TpC and GpG. These findings align with the overall expression of dinucleotide pairs across the genus, with GpC and ApC being more prevalent and TpC and GpG less so. The underrepresentation of CpG across CoPS and CoNS is noteworthy due to its potential link to immune evasion, as CpG dinucleotides are often targets for host immune recognition. This genomic trait could contribute to the pathogen's ability to establish infections without triggering a robust immune response.

Our analysis also established a significant correlation between the codon adaptation index (CAI) and G-C content, as well as between CAI and the primary axes (Axis1 and Axis2) of RSCU correspondence analysis. CoNS species, characterized by lower G-C content, demonstrated higher CAI values, suggesting greater synonymous codon bias. This relationship underscores how the *Staphylococcus* genus employs a balance between codon bias and genomic G-C content to optimize gene expression. Higher CAI values in species with lower G-C content may reflect adaptive strategies for maintaining efficient protein synthesis under specific environmental or host conditions.

Furthermore, the amino acid usage analysis across *Staphylococcus* species revealed a preference for energetically cheaper amino acids such as leucine, isoleucine, and lysine, while cysteine and methionine were less frequently used. This trend towards energy-efficient amino acid usage supports the notion of adaptive evolution that favors metabolic efficiency and cellular fitness. The correlations observed in GRAVY and AROMO analyses between amino acid hydropathicity and relative usage highlight multifactorial influences on amino acid selection and codon usage. Additionally, genomic G-C content significantly affected amino acid preferences, further demonstrating how genomic composition can influence fundamental biological processes and adaptive strategies.

In conclusion, this study provides a comprehensive overview of the codon usage patterns and their evolutionary implications within the *Staphylococcus* genus. The observed variations in G-C content, codon bias, and nucleotide preferences reflect a delicate interplay of natural selection, translational efficiency, and mutational pressures that shape the genetic landscape of these bacteria. The rigorous integration of *in vivo* and *in silico* analyses validated the reliability of the bioinformatics approaches,

ensuring accurate representation of genomic properties that impact codon preference and adaptation. These adaptive mechanisms, including the preference for A/T-rich codons, selective amino acid usage, and the underrepresentation of immune-targeting dinucleotide pairs, contribute to the pathogenic potential and survival strategies of *Staphylococcus* species. Understanding these codon usage dynamics offers valuable insights into microbial evolution and can inform future research on bacterial pathogenesis, gene expression regulation, and potential therapeutic targets.

CHAPTER 7

References

1. Huang, J., *et al*., *Genome-wide distribution and organization of microsatellites in six species of birds.* Biochemical Systematics and Ecology, 2016. **67**: p. 95-102.
2. Kovařovic, V., *et al*., *Staphylococcus ratti sp. nov. isolated from a lab rat.* Pathogens, 2022. **11**(1): p. 51.
3. Valero, A., *et al*., *Modelling the growth boundaries of Staphylococcus aureus: Effect of temperature, pH and water activity.* International Journal of Food Microbiology, 2009. **133**(1-2): p. 186-194.
4. Sato, M., *et al*., *Evaluation of culture media for Candida albicans and Staphylococcus aureus recovery in swimming pools.* Water Research, 1995. **29**(10): p. 2412-2416.
5. Nagase, N., *et al*., *Isolation and species distribution of staphylococci from animal and human skin.* Journal of Veterinary Medical Science, 2002. **64**(3): p. 245-250.
6. Scheffler, P., *et al*., *Serum thyroglobulin improves the sensitivity of the McGill Thyroid Nodule Score for well-differentiated thyroid cancer.* Thyroid, 2014. **24**(5): p. 852-857.
7. Vogel, B., *et al*., *Multivariate miRNA signatures as biomarkers for non-ischaemic systolic heart failure.* European heart journal, 2013. **34**(36): p. 2812-2823.
8. Becker, K., R.L. Skov, and C. von Eiff, *Staphylococcus, Micrococcus, and other catalase-positive cocci.* Manual of clinical microbiology, 2015: p. 354-382.
9. Baird-Parker, A., *The classification of staphylococci and micrococci from world-wide sources.* Microbiology, 1965. **38**(3): p. 363-387.
10. Shaw, C., J.M. Stitt, and S. Cowan, *Staphylococci and their classification.* Microbiology, 1951. **5**(5): p. 1010-1023.
11. Newsom, S., *MRSA and its Predecessor-A Historical View Part One: Discovery and Definition.* British Journal of Infection Control, 2003. **4**(6): p. 25-27.
12. Pal, M., K.P. Gutama, and T. Koliopoulos, *Staphylococcus aureus, an important pathogen of public health and economic importance: A comprehensive review.*

Journal of Emerging Environmental Technologies and Health Protection, 2021. **4**(2): p. 17-32.

13. Schleifer, K.H. and W.E. Kloos, *Isolation and characterization of Staphylococci from human skin I. Amended descriptions of Staphylococcus epidermidis and Staphylococcus saprophyticus and descriptions of three new species: Staphylococcus cohnii, Staphylococcus haemolyticus, and Staphylococcus xylosus.* International Journal of Systematic and Evolutionary Microbiology, 1975. **25**(1): p. 50-61.
14. Milani, M., *et al*., *Staphylococcus aureus*, in *Bacterial Degradation of Organic and Inorganic Materials: Staphylococcus aureus Meets the Nanoworld*. 2023, Springer. p. 3-20.
15. Jeffries, L., *Sensitivity to novobiocin and lysozyme in the classification of Micrococcaceae.* Journal of Applied Bacteriology, 1968. **31**(4): p. 436-442.
16. Peacock, S., *Staphylococcus aureus.* Principles and practice of clinical bacteriology, 2006. **2**: p. 73-98.
17. Rain, F.F. and A.F.M. Aslam, *The first DNA barcode of medically important cockroaches in Bangladesh.* AsPac J. Mol. Biol. Biotechnol, 2023. **31**(2): p. 80-90.
18. Abdurrahmanm, G. and B.M. Bröker, *Staphylococcus aureus and Its Proteins.* Chronic rhinosinusitis: The mucosal concept, 2022: p. 121-131.
19. Ludwig, W., *et al*., *A phylogenetic analysis of staphylococci, Peptococcus saccharolyticus and Micrococcus mucilaginosus.* Microbiology, 1981. **125**(2): p. 357-366.
20. Gherardi, G., G. Di Bonaventura, and V. Savini, *Staphylococcal taxonomy*, in *Pet-To-Man Travelling Staphylococci*. 2018, Elsevier. p. 1-10.
21. Kusuma, C.M. and J.F. Kokai-Kun, *Comparison of four methods for determining lysostaphin susceptibility of various strains of Staphylococcus aureus.* Antimicrobial agents and chemotherapy, 2005. **49**(8): p. 3256-3263.
22. Cheung, G.Y., J.S. Bae, and M. Otto, *Pathogenicity and virulence of Staphylococcus aureus.* Virulence, 2021. **12**(1): p. 547-569.
23. Oliveira, D., A. Borges, and M. Simões, *Staphylococcus aureus toxins and their molecular activity in infectious diseases.* Toxins, 2018. **10**(6): p. 252.

24. Larkin, E., *et al*., *Staphylococcus aureus: the toxic presence of a pathogen extraordinaire.* Current medicinal chemistry, 2009. **16**(30): p. 4003-4019.
25. Linage, B., *et al*., *Characterization of coagulase-positive staphylococci isolated from tank and silo ewe milk.* Journal of dairy science, 2012. **95**(4): p. 1639-1644.
26. Normanno, G., *et al*., *Coagulase-positive Staphylococci and Staphylococcus aureus in food products marketed in Italy.* International journal of food microbiology, 2005. **98**(1): p. 73-79.
27. Bond, R. and A. Loeffler, *What's happened to Staphylococcus intermedius? Taxonomic revision and emergence of multi-drug resistance.* Journal of Small Animal Practice, 2012. **53**(3): p. 147-154.
28. Kosecka-Strojek, M., A. Buda, and J. Międzobrodzki, *Staphylococcal ecology and epidemiology*, in *Pet-To-Man Travelling Staphylococci*. 2018, Elsevier. p. 11-24.
29. Chrobak-Chmiel, D., *et al*., *Staphylococcus pseudintermedius, both commensal and pathogen.* 2018.
30. Xue, L., *et al*., *Staphyloxanthin: a potential target for antivirulence therapy.* Infection and drug resistance, 2019: p. 2151-2160.
31. Blondeau, L.D., *et al*., *Zoonotic Staphylococcus pseudintermedius: an underestimated human pathogen?* Future Microbiology, 2023. **18**(6): p. 311-315.
32. Foissac, M., *et al*., *Spondylodiscitis and bacteremia due to Staphylococcus hyicus in an immunocompetent man.* Germs, 2016. **6**(3): p. 106.
33. Zhang, J., *et al*., *Genetic and virulent difference between pigmented and non-pigmented Staphylococcus aureus.* Frontiers in microbiology, 2018. **9**: p. 598.
34. Kloos, W.E. and T.L. Bannerman, *Update on clinical significance of coagulase-negative staphylococci.* Clinical microbiology reviews, 1994. **7**(1): p. 117-140.
35. Pyörälä, S. and S. Taponen, *Coagulase-negative staphylococci—Emerging mastitis pathogens.* Veterinary microbiology, 2009. **134**(1-2): p. 3-8.
36. Rogers, K.L., P.D. Fey, and M.E. Rupp, *Coagulase-negative staphylococcal infections.* Infectious disease clinics of North America, 2009. **23**(1): p. 73-98.

37. Bal, E.B.B., S. Bayar, and M.A. Bal, *Antimicrobial susceptibilities of coagulase-negative staphylococci (CNS) and streptococci from bovine subclinical mastitis cases.* The Journal of Microbiology, 2010. **48**: p. 267-274.
38. Cunha, M.d.L.R., Y.K. Sinzato, and L.V. Silveira, *Comparison of methods for the identification of coagulase-negative staphylococci.* Memórias do Instituto Oswaldo Cruz, 2004. **99**: p. 855-860.
39. Lee, A.S., *et al*., *Methicillin-resistant Staphylococcus aureus.* Nature reviews Disease primers, 2018. **4**(1): p. 1-23.
40. Nakatsuji, T., *et al*., *Antimicrobials from human skin commensal bacteria protect against Staphylococcus aureus and are deficient in atopic dermatitis.* Science translational medicine, 2017. **9**(378): p. eaah4680.
41. Tenover, F.C. and R.J. Gorwitz, *The epidemiology of Staphylococcus infections.* Gram-positive pathogens, 2006: p. 526-534.
42. MUSHER, D.M. and S.O. McKENZIE, *Infections due to Staphylococcus aureus.* Medicine, 1977. **56**(5): p. 383-410.
43. Becker, K., *Pathogenesis of Staphylococcus aureus*, in *Staphylococcus aureus*. 2018, Elsevier. p. 13-38.
44. Laurent, F. and M. Butin, *Staphylococcus capitis and NRCS-A clone: the story of an unrecognized pathogen in neonatal intensive care units.* Clinical Microbiology and Infection, 2019. **25**(9): p. 1081-1085.
45. Heath, V., *et al*., *Staphylococcus capitis: review of its role in infections and outbreaks.* Antibiotics, 2023. **12**(4): p. 669.
46. Moroni, P., *et al*., *Subclinical mastitis and antimicrobial susceptibility of Staphylococcus caprae and Staphylococcus epidermidis isolated from two Italian goat herds.* Journal of Dairy Science, 2005. **88**(5): p. 1694-1704.
47. Gowda, A., A.L. Pensiero, and C.D. Packer, *Staphylococcus caprae: a skin commensal with pathogenic potential.* Cureus, 2018. **10**(10).
48. Bückle, A., *et al*., *Genetic diversity and population structure of food-borne Staphylococcus carnosus strains.* Systematic and Applied Microbiology, 2017. **40**(1): p. 34-41.

49. Stefano, M., *et al*., *Intrascrotal Abscess, Propionibacterium acnes and Staphylococcus cohnii ssp. cohnii: A Case Report and Review of the Literature.* Case Reports in Urology, 2012. **2012**(1): p. 313694.
50. Szewczyk, E.M., *et al*., *Potential role of Staphylococcus cohnii in a hospital environment.* Microbial ecology in health and disease, 2003. **15**(1): p. 51-56.
51. Gharsa, H., *et al*., *Characterisation of nasal S taphylococcus delphini and S taphylococcus pseudintermedius isolates from healthy donkeys in T unisia.* Equine veterinary journal, 2015. **47**(4): p. 463-466.
52. Brown, M.M. and A.R. Horswill, *Staphylococcus epidermidis—Skin friend or foe?* PLoS pathogens, 2020. **16**(11): p. e1009026.
53. Jeong, D.-W., *et al*., *A proposal to unify two subspecies of Staphylococcus equorum: Staphylococcus equorum subsp. equorum and Staphylococcus equorum subsp. linens.* Antonie Van Leeuwenhoek, 2013. **104**: p. 1049-1062.
54. Devriese*, L., *et al*., *Staphylococcus gallinarum and Staphylococcus caprae, two new species from animals.* International Journal of Systematic and Evolutionary Microbiology, 1983. **33**(3): p. 480-486.
55. Sousa, M., *et al*., *Antimicrobial resistance determinants in Staphylococcus spp. recovered from birds of prey in Portugal.* Veterinary Microbiology, 2014. **171**(3-4): p. 436-440.
56. Wolden, R., *et al*., *Identification of surface proteins in a clinical Staphylococcus haemolyticus isolate by bacterial surface shaving.* BMC microbiology, 2020. **20**: p. 1-18.
57. Wolden, R., *Adhesion mechanisms and bacteriocins in Staphylococcus haemolyticus-New targets for the prevention and treatment of infections.* 2024.
58. Severn, M.M., *et al*., *The ubiquitous human skin commensal Staphylococcus hominis protects against opportunistic pathogens.* MBio, 2022. **13**(3): p. e00930-22.
59. Frickmann, H., *et al*., *Comparison of the etiological relevance of Staphylococcus haemolyticus and Staphylococcus hominis.* European Journal of Clinical Microbiology & Infectious Diseases, 2018. **37**: p. 1539-1545.

60. Ross Fitzgerald, J., *The Staphylococcus intermedius group of bacterial pathogens: species re-classification, pathogenesis and the emergence of meticillin resistance.* Veterinary dermatology, 2009. **20**(5-6): p. 490-495.
61. Kamarudin, K.R., *et al*., *Isolation of a Pigment-producing Strain of Staphylococcus kloosii from the Respiratory Tree of Holothuria (Mertensiothuria) leucospilota () from Malaysian Waters.* Tropical Life Sciences Research, 2013. **24**(1): p. 85.
62. Rivera, M., *et al*., *Staphylococcus lentus peritonitis: a case report.* Peritoneal dialysis international: journal of the International Society for Peritoneal Dialysis, 2014. **34**(4): p. 469.
63. Hovelius, B. and P.-A. Mårdh, *Staphylococcus saprophyticus as a common cause of urinary tract infections.* Reviews of infectious diseases, 1984. **6**(3): p. 328-337.
64. Nemeghaire, S., *et al*., *The ecological importance of the Staphylococcus sciuri species group as a reservoir for resistance and virulence genes.* Veterinary Microbiology, 2014. **171**(3-4): p. 342-356.
65. Shields, B.E., A.J. Tschetter, and K.A. Wanat, *Staphylococcus simulans: An emerging cutaneous pathogen.* JAAD case reports, 2016. **2**(6): p. 428-429.
66. Webster, J.A., *et al*., *Identification of the Staphylococcus sciuri Species Group with Eco RI Fragments Containing rRNA Sequences and Description of Staphylococcus vitulus sp. nov.* International Journal of Systematic and Evolutionary Microbiology, 1994. **44**(3): p. 454-460.
67. Ravaioli, S., *et al*., *The Opportunistic Pathogen Staphylococcus warneri: Virulence and Antibiotic Resistance, Clinical Features, Association with Orthopedic Implants and Other Medical Devices, and a Glance at Industrial Applications.* Antibiotics, 2024. **13**(10): p. 972.
68. Dordet-Frisoni, E., *et al*., *Genomic diversity in Staphylococcus xylosus.* Applied and environmental microbiology, 2007. **73**(22): p. 7199-7209.
69. Paul, P., A.K. Malakar, and S. Chakraborty, *Codon usage and amino acid usage influence genes expression level.* Genetica, 2018. **146**: p. 53-63.
70. Parvathy, S.T., V. Udayasuriyan, and V. Bhadana, *Codon usage bias.* Molecular biology reports, 2022. **49**(1): p. 539-565.

71. Sharp, P.M., *et al*., *Variation in the strength of selected codon usage bias among bacteria.* Nucleic acids research, 2005. **33**(4): p. 1141-1153.
72. Sau, K., *et al*., *Synonymous codon usage bias in 16 Staphylococcus aureus phages: implication in phage therapy.* Virus research, 2005. **113**(2): p. 123-131.
73. Ran, W. and P.G. Higgs, *The influence of anticodon–codon interactions and modified bases on codon usage bias in bacteria.* Molecular biology and evolution, 2010. **27**(9): p. 2129-2140.
74. Behura, S.K. and D.W. Severson, *Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes.* Biological Reviews, 2013. **88**(1): p. 49-61.
75. Tekaia, F., E. Yeramian, and B. Dujon, *Amino acid composition of genomes, lifestyles of organisms, and evolutionary trends: a global picture with correspondence analysis.* Gene, 2002. **297**(1-2): p. 51-60.
76. Shah, P. and M.A. Gilchrist, *Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift.* Proceedings of the National Academy of Sciences, 2011. **108**(25): p. 10231-10236.
77. Rocha, E.P., *Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization.* Genome research, 2004. **14**(11): p. 2279-2286.
78. Muyle, A., *et al*., *GC-biased gene conversion and selection affect GC content in the Oryza genus (rice).* Molecular biology and evolution, 2011. **28**(9): p. 2695-2706.
79. Plotkin, J.B. and G. Kudla, *Synonymous but not the same: the causes and consequences of codon bias.* Nature Reviews Genetics, 2011. **12**(1): p. 32-42.
80. Sahoo, S. and S. Das, *Analyzing gene expression and codon usage bias in diverse genomes using a variety of models.* Current Bioinformatics, 2014. **9**(2): p. 102-112.
81. Vijaranakul, U., *et al*., *Cloning and nucleotide sequencing of a Staphylococcus aureus gene encoding a branched-chain-amino-acid transporter.* Applied and environmental microbiology, 1998. **64**(2): p. 763-767.

82. Mahfooz, S., *et al*., *Comparative genomics reveals genus specific encoding of amino acids by tri-nucleotide SSRs in human pathogenic Streptococcus and Staphylococcus bacteria.* Biologia, 2022. **77**(10): p. 2955-2966.

83. Votintseva, A.A., *et al*., *Prevalence of Staphylococcus aureus protein A (spa) mutants in the community and hospitals in Oxfordshire.* BMC microbiology, 2014. **14**: p. 1-11.

84. Clegg, J., *et al*., *Staphylococcus aureus vaccine research and development: the past, present and future, including novel therapeutic strategies.* Frontiers in immunology, 2021. **12**: p. 705360.

85. Melles, D.C., *et al*., *Panton-Valentine leukocidin genes in Staphylococcus aureus.* Emerging infectious diseases, 2006. **12**(7): p. 1174.

86. Angov, E., *Codon usage: nature's roadmap to expression and folding of proteins.* Biotechnology journal, 2011. **6**(6): p. 650-659.

87. Ge, Z., *et al*., *Viral adaption of staphylococcal phage: A genome-based analysis of the selective preference based on codon usage Bias.* Genomics, 2020. **112**(6): p. 4657-4665.

88. Roy, A., *et al*., *Comparative investigation of the various determinants that influence the codon and amino acid usage patterns in the genus Bifidobacterium.* World Journal of Microbiology and Biotechnology, 2015. **31**: p. 959-981.

89. Naushad, S., *et al*., *Comprehensive phylogenetic analysis of bovine non-aureus staphylococci species based on whole-genome sequencing.* Frontiers in Microbiology, 2016. **7**: p. 235412.

90. Friedman, R.C., *et al*., *Common and phylogenetically widespread coding for peptides by bacterial small RNAs.* BMC genomics, 2017. **18**: p. 1-21.

91. Sharma, A., S. Gupta, and K. Paul, *Codon usage behavior distinguishes pathogenic Clostridium species from the non-pathogenic species.* Gene, 2023. **873**: p. 147394.

92. Liadaki, K., *et al*., *PDE4 Gene Family Variants Are Associated with Response to Apremilast Treatment in Psoriasis.* Genes, 2024. **15**(3): p. 369.

93. Iriarte, A., G. Lamolle, and H. Musto, *Codon usage bias: an endless tale.* Journal of molecular evolution, 2021. **89**(9): p. 589-593.

94. Bahiri-Elitzur, S. and T. Tuller, *Codon-based indices for modeling gene expression and transcript evolution.* Computational and Structural Biotechnology Journal, 2021. **19**: p. 2646-2663.

95. Ostash, B. and M. Anisimova, *Visualizing codon usage within and across genomes: concepts and tools.* Statistical Modelling and Machine Learning Principles for Bioinformatics Techniques, Tools, and Applications, 2020: p. 213-288.

96. Novoa, E.M., *et al*., *Elucidation of codon usage signatures across the domains of life.* Molecular Biology and Evolution, 2019. **36**(10): p. 2328-2339.

97. Hart, A., *et al*., *Codon usage bias reveals genomic adaptations to environmental conditions in an acidophilic consortium.* PloS one, 2018. **13**(5): p. e0195869.

98. Arella, D., M. Dilucca, and A. Giansanti, *Codon usage bias and environmental adaptation in microbial organisms.* Molecular Genetics and Genomics, 2021. **296**(3): p. 751-762.

99. Fu, Y., *et al*., *Codon usage bias analysis in macronuclear genomes of ciliated protozoa.* Microorganisms, 2023. **11**(7): p. 1833.

100. Smith, R.D., *Enhanced effective codon numbers to understand codon usage bias.* Biosystems, 2022. **220**: p. 104734.

101. Li, G., L. Zhang, and P. Xue, *Codon usage pattern and genetic diversity in chloroplast genomes of Panicum species.* Gene, 2021. **802**: p. 145866.

102. Xu-Yuan Liua, Y., Kai-Kai Jia, Jie Zhua, Peng Linga, Tao Zhoua, Lan-Ying Fanb, Shang-Qian Xiea,, *Genome-wide codon usage pattern analysis reveals the correlation between*

codon usage bias and gene expression in Cuscuta australis. Genomics, 2020. **112**: p. 2695-2702.

103. Zhang, P., *et al*., *Analysis of codon usage bias of chloroplast genomes in Gynostemma species.* Physiology and Molecular Biology of Plants, 2021: p. 1-11.

104. Chakraborty, S., S. Yengkhom, and A. Uddin, *Analysis of codon usage bias of chloroplast genes in Oryza species: Codon usage of chloroplast genes in Oryza species.* Planta, 2020. **252**: p. 1-20.

105. Anwar, A.M., M. Aljabri, and M. El-Soda, *Patterns of genome-wide codon usage bias in tobacco, tomato and potato.* Biotechnology & Biotechnological Equipment, 2021. **35**(1): p. 657-664.

106. Gao, Y., *et al*., *Analysis of codon usage bias of WRKY transcription factors in Helianthus annuus.* BMC Genomic Data, 2022. **23**(1): p. 46.

107. Xu, Y., *et al*., *Codon usage bias regulates gene expression and protein conformation in yeast expression system P. pastoris.* Microbial cell factories, 2021. **20**(1): p. 91.

108. Wang, F., *et al*., *Codon usage bias analysis of mitochondrial protein-coding genes in 12 species of Candida.* Journal of Genetics, 2023. **102**(2): p. 36.

109. Baeza, M., *et al*., *Codon usage bias in yeasts and its correlation with gene expression, growth temperature, and protein structure.* Frontiers in Microbiology, 2024. **15**: p. 1414422.

110. Liu, Y., Q. Yang, and F. Zhao, *Synonymous but not silent: the codon usage code for gene expression and protein folding.* Annual review of biochemistry, 2021. **90**(1): p. 375-401.

111. Deng, Y., *et al*., *Hidden patterns of codon usage bias across kingdoms.* Journal of The Royal Society Interface, 2020. **17**(163): p. 20190819.

112. Tang, D., *et al*., *Analysis of codon usage bias and evolution in the chloroplast genome of Mesona chinensis Benth.* Development Genes and Evolution, 2021. **231**: p. 1-9.

113. Khandia, R., *et al*., *Codon usage bias correlates with gene length in neurodegeneration associated genes.* Frontiers in Neuroscience, 2022. **16**: p. 895607.

114. Zhao, F., *et al*., *Genome-wide role of codon usage on transcription and identification of potential regulators.* Proceedings of the National Academy of Sciences, 2021. **118**(6): p. e2022590118.

115. Zhao, Y., *et al*., *Codon Usage Bias: A Potential Factor Affecting VGLUT Developmental Expression and Protein Evolution.* Molecular Neurobiology, 2024: p. 1-15.

116. Kumar, N., *et al*., *Evolutionary signatures governing the codon usage bias in coronaviruses and their implications for viruses infecting various bat species.* Viruses, 2021. **13**(9): p. 1847.

117. Wang, H., *et al*., *Codon usage bias of Venezuelan equine encephalitis virus and its host adaption.* Virus research, 2023. **328**: p. 199081.

118. Hussain, S., *et al*., *Analysis of codon usage and nucleotide bias in middle east respiratory syndrome coronavirus genes.* Evolutionary Bioinformatics, 2020. **16**: p. 1176934320918861.

119. Ghorbani, A., *Genetic analysis of tomato brown rugose fruit virus reveals evolutionary adaptation and codon usage bias patterns.* Scientific Reports, 2024. **14**(1): p. 21281.

120. Deb, B., A. Uddin, and S. Chakraborty, *Codon usage pattern and its influencing factors in different genomes of hepadnaviruses.* Archives of virology, 2020. **165**: p. 557-570.

121. Srivastava, S., *et al*., *Codon Usage Biasness-Cause, Consequences and Implications in Evolutionary Development.* Biotech Today, 2019. **9**(2): p. 8-17.

122. Wright, F., *The 'effective number of codons' used in a gene.* Gene, 1990. **87**(1): p. 23-29.

123. Xu, C., *et al*., *Analysis of synonymous codon usage patterns in seven different citrus species.* Evolutionary Bioinformatics, 2013. **9**: p. EBO. S11930.

124. Liu, Y.-s., *et al*., *Analysis of synonymous codon usage in porcine reproductive and respiratory syndrome virus.* Infection, Genetics and Evolution, 2010. **10**(6): p. 797-803.

125. Wang, L., *et al*., *Genome-wide analysis of codon usage bias in four sequenced cotton species.* PloS one, 2018. **13**(3): p. e0194372.

126. Deb, B., *et al*., *Analysis of codon usage pattern of mitochondrial protein-coding genes in different hookworms.* Molecular and biochemical parasitology, 2018. **219**: p. 24-32.

127. Morla, S., A. Makhija, and S. Kumar, *Synonymous codon usage pattern in glycoprotein gene of rabies virus.* Gene, 2016. **584**(1): p. 1-6.

128. Andargie, M. and Z. Congyi, *Genome-wide analysis of codon usage in sesame (Sesamum indicum L.).* Heliyon, 2022. **8**(1).

129. Guan, D.-L., *et al*., *Analysis of codon usage patterns in Hirudinaria manillensis reveals a preference for GC-ending codons caused by dominant selection constraints.* BMC genomics, 2018. **19**: p. 1-14.

130. Chen, Y., *A comparison of synonymous codon usage bias patterns in DNA and RNA virus genomes: quantifying the relative importance of mutational pressure and natural selection.* BioMed research international, 2013. **2013**.

131. Roy, A. and J. van Staden, *Insights into the riddles of codon usage patterns and codon context signatures in fungal genus Puccinia, a persistent threat to global agriculture.* Journal of Cellular Biochemistry, 2019. **120**(12): p. 19555-19566.

132. Roy, A. and J. van Staden, *Comprehensive profiling of codon usage signatures and codon context variations in the genus Ustilago.* World Journal of Microbiology and Biotechnology, 2019. **35**: p. 1-15.

133. Gupta, S., K. Paul, and A. Roy, *Codon usage signatures in the genus Cryptococcus: A complex interplay of gene expression, translational selection and compositional bias.* Genomics, 2021. **113**(1): p. 821-830.

134. Khandia, R., T. Alqahtani, and A.M. Alqahtani, *Genes common in primary immunodeficiencies and cancer display overrepresentation of codon CTG and dominant role of selection pressure in shaping codon usage.* Biomedicines, 2021. **9**(8): p. 1001.

135. Chen, Y., *et al*., *Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Senecavirus A evolution.* Microbial pathogenesis, 2017. **112**: p. 313-319.

136. Nyayanit, D.A., *et al*., *Natural selection plays an important role in shaping the codon usage of structural genes of the viruses belonging to the coronaviridae family.* Viruses, 2020. **13**(1): p. 3.

137. Chakraborty, S., *et al*., *Analysis of codon usage patterns and influencing factors in Nipah virus.* Virus research, 2019. **263**: p. 129-138.

138. Hanson, G. and J. Coller, *Codon optimality, bias and usage in translation and mRNA decay.* Nature reviews Molecular cell biology, 2018. **19**(1): p. 20-30.

139. Puigbò, P., I.G. Bravo, and S. Garcia-Vallvé, *E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI).* BMC bioinformatics, 2008. **9**: p. 1-7.

140. Carbone, A., A. Zinovyev, and F. Képes, *Codon adaptation index as a measure of dominating codon bias.* Bioinformatics, 2003. **19**(16): p. 2005-2015.
141. Sablok, G., *et al*., *Synonymous codon usage, GC 3, and evolutionary patterns across plastomes of three pooid model species: Emerging grass genome models for monocots.* Molecular biotechnology, 2011. **49**: p. 116-128.
142. Saha, J., S. Dey, and A. Pal, *Whole genome sequencing and comparative genomic analyses of Pseudomonas aeruginosa strain isolated from arable soil reveal novel insights into heavy metal resistance and codon biology.* Current Genetics, 2022. **68**(3-4): p. 481-503.
143. Baran, R.H. and H. Ko, *Detecting horizontally transferred and essential genes based on dinucleotide relative abundance.* DNA research, 2008. **15**(5): p. 267-276.
144. Tao, P., *et al*., *Analysis of synonymous codon usage in classical swine fever virus.* Virus genes, 2009. **38**: p. 104-112.
145. Shen, Z., *et al*., *Analysis of codon usage patterns in citrus based on coding sequence data.* BMC genomics, 2020. **21**: p. 1-10.
146. Zhang, Y., *et al*., *Codon usage patterns across seven Rosales species.* BMC Plant Biology, 2022. **22**(1): p. 65.
147. Kunec, D. and N. Osterrieder, *Codon pair bias is a direct consequence of dinucleotide bias.* Cell reports, 2016. **14**(1): p. 55-67.
148. Şen, A., *et al*., *Codon optimization: a mathematical programing approach.* Bioinformatics, 2020. **36**(13): p. 4012-4020.
149. Chakraborty, S., *et al*., *Codon usage pattern and prediction of gene expression level in Bungarus species.* Gene, 2017. **604**: p. 48-60.
150. George, D. and P. Mallery, *IBM SPSS statistics 26 step by step: A simple guide and reference*. 2019: Routledge.
151. Long, S., *et al*., *Analysis of compositional bias and codon usage pattern of the coding sequence in Banna virus genome.* Virus research, 2018. **258**: p. 68-72.
152. McKnight, D.T., *et al*., *microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies.* Environmental DNA, 2019. **1**(1): p. 14-25.

153. Wood, D.E. and S.L. Salzberg, *Kraken: ultrafast metagenomic sequence classification using exact alignments.* Genome biology, 2014. **15**: p. 1-12.

154. Andrews, S., *FastQC: a quality control tool for high throughput sequence data. 2010*. 2017.

155. Shaw, J. and Y.W. Yu, *Fast and robust metagenomic sequence comparison through sparse chaining with skani.* Nature Methods, 2023. **20**(11): p. 1661-1665.

156. Yoon, S.-H., *et al*., *Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies.* International journal of systematic and evolutionary microbiology, 2017. **67**(5): p. 1613-1617.

157. Lee, I., *et al*., *ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences.* International journal of systematic and evolutionary microbiology, 2017. **67**(6): p. 2053-2057.

158. De Nies, L., *et al*., *PathoFact: a pipeline for the prediction of virulence factors and antimicrobial resistance genes in metagenomic data.* Microbiome, 2021. **9**: p. 1-14.

159. Alcock, B.P., *et al*., *CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database.* Nucleic acids research, 2023. **51**(D1): p. D690-D699.

160. Meier-Kolthoff, J.P. and M. Göker, *TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy.* Nature communications, 2019. **10**(1): p. 2182.

161. Brocchieri, L., *The GC content of bacterial genomes.* J Phylogenetics Evol Biol, 2014. **2**: p. 1-3.

162. Rocha, E.P. and A. Danchin, *Base composition bias might result from competition for metabolic resources.* TRENDS in Genetics, 2002. **18**(6): p. 291-294.

163. Nasrullah, I., *et al*., *Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution.* BMC evolutionary biology, 2015. **15**: p. 1-15.

164. Duan, J. and M.A. Antezana, *Mammalian mutation pressure, synonymous codon choice, and mRNA degradation.* Journal of Molecular Evolution, 2003. **57**: p. 694-701.

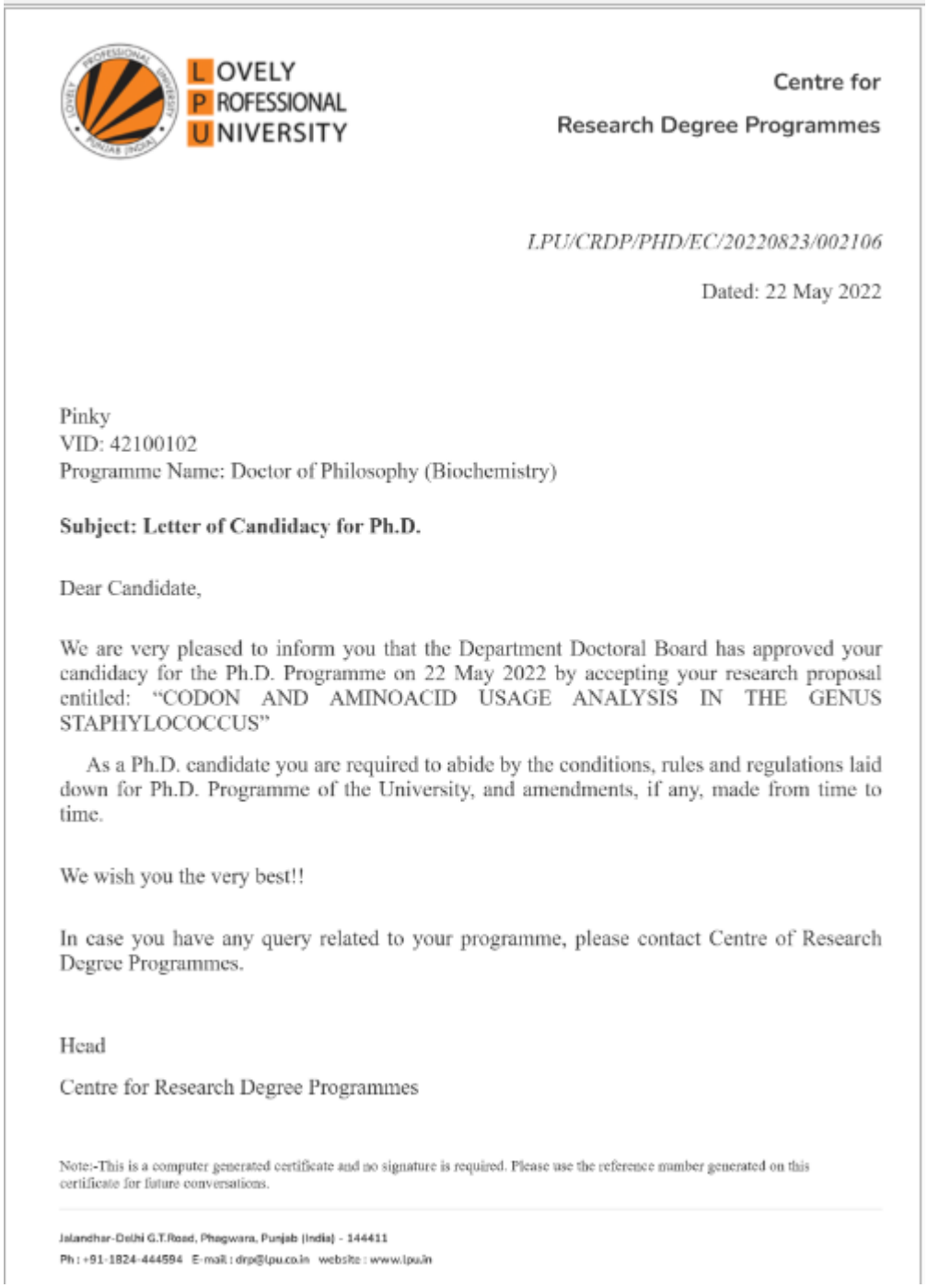
165. Kalia, D., *et al*., *Nucleotide, c-di-GMP, c-di-AMP, cGMP, cAMP,(p) ppGpp signaling in bacteria and implications in pathogenesis.* Chemical Society Reviews, 2013. **42**(1): p. 305-341.

166. Kuznik, A., G. Panter, and R. Jerala, *Recognition of nucleic acids by Toll-like receptors and development of immunomodulatory drugs.* Current medicinal chemistry, 2010. **17**(18): p. 1899-1914.

167. Shirota, H. and D. Klinman, *CpG Oligodeoxynucleotides as adjuvants for clinical use.* Immunopotentiators in modern vaccines, 2017: p. 163-198.

APPENDIX

I. Letter of Candidacy



LOVELY PROFESSIONAL UNIVERSITY

Centre for
Research Degree Programmes

LPU/CRDP/PHD/EC/20220823/002106

Dated: 22 May 2022

Pinky
VID: 42100102
Programme Name: Doctor of Philosophy (Biochemistry)

Subject: Letter of Candidacy for Ph.D.

Dear Candidate,

We are very pleased to inform you that the Department Doctoral Board has approved your candidacy for the Ph.D. Programme on 22 May 2022 by accepting your research proposal entitled: "CODON AND AMINOACID USAGE ANALYSIS IN THE GENUS STAPHYLOCOCCUS"

As a Ph.D. candidate you are required to abide by the conditions, rules and regulations laid down for Ph.D. Programme of the University, and amendments, if any, made from time to time.

We wish you the very best!!

In case you have any query related to your programme, please contact Centre of Research Degree Programmes.

Head

Centre for Research Degree Programmes

Note:-This is a computer generated certificate and no signature is required. Please use the reference number generated on this certificate for future conversations.

Jalandhar-Delhi G.T.Road, Phagwara, Punjab (India) - 144411
Ph : +91-1824-444594 E-mail : drp@lpu.co.in website : www.lpu.in

II. List of Publications

1. **Arora, P**., Mukhopadhyay, C. S., & Kaur, S. (2024). Comparative genome-wide analysis of codon usage of *Staphylococcus* genus. *Current Genetics, 70*(1), 10.

Current Genetics (2024) 70:10
https://doi.org/10.1007/s00294-024-01297-3

ORIGINAL ARTICLE

Comparative genome wise analysis of codon usage of *Staphylococcus* Genus

Pinky Arora[1] · Chandra Shekhar Mukhopadhyay[2] · Sandeep Kaur[3]

Received: 29 March 2024 / Revised: 5 July 2024 / Accepted: 22 July 2024

Abstract
The genus *Staphylococcus* encompasses a diverse array of bacteria with significant implications for human health, including disreputable pathogens such as *Staphylococcus aureus* and *Staphylococcus epidermidis*. Understanding the genetic composition and codon usage patterns of *Staphylococcus* species is crucial for unraveling their evolutionary dynamics, adaptive strategies, and pathogenic potential. In this study, we conducted a comprehensive analysis of codon usage patterns across 48 species within the *Staphylococcus* genus. Our findings uncovered variations in genomic G-C content across Staphylococcus species, impacting codon usage preferences, with a notable preference for A/T-rich codons observed in pathogenic strains. This preference for A/T-rich codons suggests an energy-saving strategy in pathogenic organisms. Analysis of dinucleotide pair expression patterns unveiled insights into genomic dynamics, with overrepresented codon pairs reflecting trends in dinucleotide expression across genomes. Additionally, a significant correlation between CAI and genomic G-C content underscored the intricate relationship between codon usage patterns and gene expression strategies. Amino acid usage analysis highlighted preferences for energetically cheaper amino acids, suggesting adaptive strategies promoting energy efficiency. This comprehensive analysis sheds light on the evolutionary dynamics and adaptive mechanisms employed by *Staphylococcus* species, providing valuable insights into their pathogenic potential and clinical implications. Understanding these genomic features is crucial for devising strategies to combat staphylococcal infections and improve public health outcomes.

Keywords *Staphylococcus* · Codon usage · Translational selection · Codon adaptive index · RSCU · ENc

Introduction

The genus *Staphylococcus*, belonging to the Micrococcaceae family, comprises gram-positive, facultative anaerobic bacteria that form clusters and display a cocci shape (Becker et al. 2015). Initially identified by Sir Alexander Ogston in 1880 from pus in a knee joint surgical abscess, Staphylococcal organisms have since been recognized as potential pathogens causing severe staph infections (Adhikari 2021). Staphylococci represent a cluster of Gram-positive cocci with a diameter ranging from 0.5 to 1 µm (Winstel et al. 2021). While they may appear singly, in pairs, or in tetrads, they most commonly form clusters resembling grapes. These microorganisms lack motility and spore-forming abilities, and they are both aerobic and facultatively anaerobic (Logan et al. 2011). These bacteria are commonly present as commensal organisms on the skin and mucous membranes of birds and warm-blooded mammals (Rubinstein 2008). Additionally, they can be found in the air, dust, water, and food particles, indicating their ubiquitous presence. *Staphylococcus* bacteria can cause a wide range of infections, including skin and bone infections (Soumya et al. 2017), bacteraemia (bloodstream infection), and endocarditis (heart valve infection) (Ferguson 1946). These bacteria can also lead to food poisoning, pneumonia, and toxic shock syndrome (TSS), a serious and potentially

✉ Sandeep Kaur
sandeep.28639@lpu.co.in

[1] School of Bioengineering and Biosciences, Lovely Professional University, Jalandhar-Delhi G.T. Road, Phagwara, Punjab 144411, India

[2] Department of Bioinformatics, College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ferozepur G.T. Road, Ludhiana, Punjab 141004, India

[3] Department of Medical Laboratory Sciences, Lovely Professional University, Phagwara, Punjab 144411, India

Published online: 31 July 2024

2. **Arora, P**., Kumar, S., & Kaur, S. (2024). Amino acid usage in coagulase-positive staphylococcal bacteria. *E3S Web of Conferences, 556*, 01054.

E3S Web of Conferences **556**, 01054 (2024) https://doi.org/10.1051/e3sconf/202455601054
RAWMU-2024

Amino acid usage in coagulase positive staphylococcal bacteria

Pinky Arora[1], *Shubham* Kumar[2], *Sandeep* Kaur[3]*

[1]School of Bioengineering and Biosciences, Lovely Professional University, Jalandhar-Delhi G.T. Road, Phagwara, Punjab 144411, India.
[2]School of Pharmaceutical Sciences, Lovely Professional University, Jalandhar-Delhi G.T. Road, Phagwara, Punjab 144411, India.
[3]Department of Medical Laboratory Sciences, Lovely Professional University, Phagwara, 144411, Punjab, India.

*Corresponding Author: sndp4482@gmail.com

Abstract: Coagulase-positive Staphylococcus (CoPS) species inhabit human and animal skin and nasal flora, posing opportunistic infection risks. This study explores amino acid preferences within CoPS, highlighting biosynthetic efficiency and functional demands in genome evolution. Leucine, isoleucine, and lysine were favored, while cysteine, methionine, proline, histidine, and tryptophan were avoided, reflecting a strategy to minimize biosynthetic costs and enhance cellular fitness. GRAVY and AROMO analyses demonstrated significant correlations between amino acid usage and protein properties like hydrophobicity and aromaticity. GRAVY scores, indicative of protein hydrophobicity, correlated positively with GC content, particularly at the third codon position ($GC3_s$), and overall GC content. Aromaticity showed similar positive correlations with GC content and hydrophobicity, suggesting a preference for aromatic amino acids in GC-rich CoPS genomes. Species-specific findings in *S. aureus* revealed correlations between $GC3_s$, hydrophobicity, and codon usage adaptation, emphasizing adaptive strategies to optimize protein stability. This study enhances understanding of CoPS evolutionary pressures and informs potential therapeutic strategies.

Keywords: Staphylococcus, Coagulase Positive Species, RAAU, GRAVY, AROMO

1 Introduction

Coagulase-positive Staphylococcus (CoPS) species are typically commensal organisms found in the skin and nasal flora of both humans and animals, yet they possess the potential to cause opportunistic infections. Notable members of this group include *Staphylococcus aureus*, *Staphylococcus pseudintermedius*, *Staphylococcus intermedius*, *Staphylococcus delphini*, *Staphylococcus lutrae*, and *Staphylococcus argenteus*, each contributing

III. Other Allied Publications

- Singh, J., Sharma, M., Singh, H., **Arora, P**., Utreja, P., & Kumar, S. (2024). Formulation, characterization and in vitro evaluation of mesalamine and *Bifidobacterium bifidum* loaded hydrogel beads in capsule system for colon targeted drug delivery. *AAPS PharmSciTech, 25*(3), 61.
- Kumar, S., Arora, P., **Wadhwa, P**., & Kaur, P. (2023). A rationalized approach to design and discover novel non-steroidal derivatives through computational aid for the treatment of prostate cancer. *Current Computer-Aided Drug Design*. https://doi.org/10.2174/1573409919666230626113346
- Gupta, S., Jha, S., Rani, S., **Arora, P**., & Kumar, S. (2024). Medicinal perspective of 2,4-thiazolidinediones derivatives: An insight into recent advancements. *ChemistryOpen*. https://doi.org/e202400147
- Gyampoh, S., Nabi, M., **Arora, P**., & Kumar, S. (2024). Exploring 5-benzylthiazolidine-2,4-dione derivatives as promising PPAR-gamma agonists through computational methods. *Current Signal Transduction Therapy*. https://doi.org/10.2174/0115743624309015240827103538
- Mwamafupa, A., **Arora, P**., Singh, J., Seksaria, K., & Kumar, S. (2024). Discovery of β-carboline-based derivatives through computational aid for the treatment of *Leishmania*. *Current Signal Transduction Therapy*, *19*(1), 29–47.

IV. Conferences Attended

ICRAHS
INTERNATIONAL CONFERENCE ON RECENT ADVANCES IN HEALTH SCIENCES

SCHOOL OF PHARMACEUTICAL SCIENCES

Certificate No. 333715

NAAC A++

Certificate of Participation

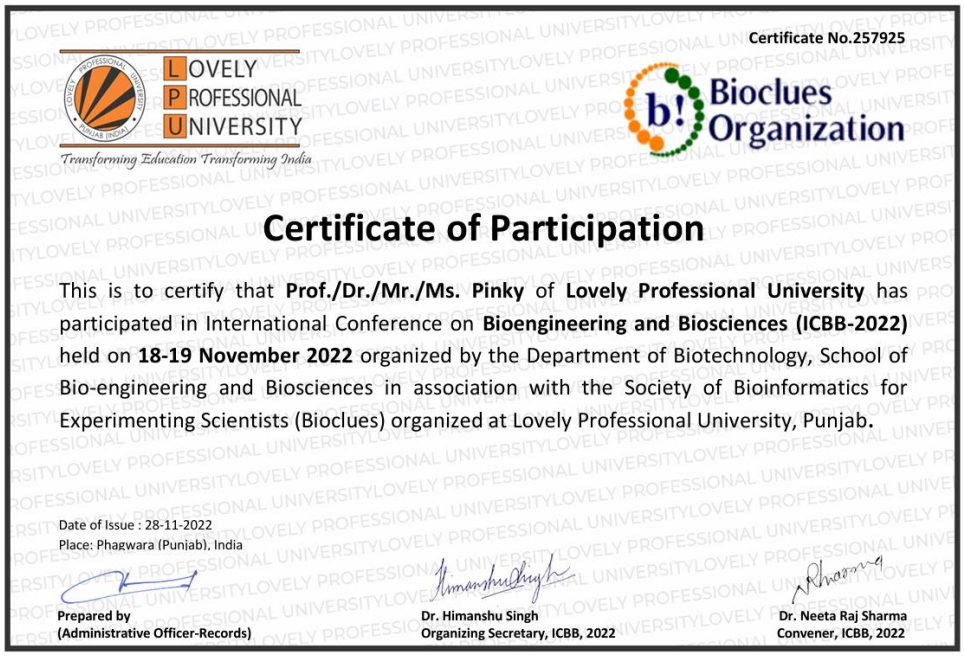
This is to certify that ***Prof./Dr./Mr./Ms. Pinky*** *has successfully participated as Delegate and Oral Presenter on topic* ***Codon usage analysis in staphylococcus genus*** *in offline mode at* ***"International Conference on Recent Advances in Health Sciences" (ICRAHS-2024)*** *under the Theme* ***"Innovations in Sustainable Health Sciences and Research"*** *held on* ***27th April 2024****, organized by School of Pharmaceutical Sciences, Lovely Professional University, Punjab.*

Date of Issue : 29-05-2024
Place : Phagwara (Punjab), India

Prepared by
(Administrative Officer-Records)

Organizing Secretory
Dr. Vikas Sharma
School of Pharmaceutical Science

Head of School
Dr. Monica Gulati
School of Pharmaceutical Science



Certificate No.257925

LOVELY PROFESSIONAL UNIVERSITY
Transforming Education Transforming India

Bioclues Organization

Certificate of Participation

This is to certify that **Prof./Dr./Mr./Ms. Pinky** of **Lovely Professional University** has participated in International Conference on **Bioengineering and Biosciences (ICBB-2022)** held on **18-19 November 2022** organized by the Department of Biotechnology, School of Bio-engineering and Biosciences in association with the Society of Bioinformatics for Experimenting Scientists (Bioclues) organized at Lovely Professional University, Punjab.

Date of Issue : 28-11-2022
Place: Phagwara (Punjab), India

Prepared by
(Administrative Officer-Records)

Dr. Himanshu Singh
Organizing Secretary, ICBB, 2022

Dr. Neeta Raj Sharma
Convener, ICBB, 2022