Research Article

Phylogeny Influences Genome Size and GC but not Sequence and Organismal Complexity in *Staphylococci*

Akash Ajay

School of Environmental Sciences, Jawaharlal Nehru University, New Delhi, INDIA.

Submission Date: 14-09-2023; Revision Date: 15-10-2023; Accepted Date: 13-11-2023.

ABSTRACT

Aim: Like other ecological and morphological features, genomic features are adaptive and can be influenced by phylogeny. While some features like genome size and genomic GC have been explored in the past, still some features like genomic repeat fraction and protein-coding genes are unexplored. Understanding the trait evolution of the individual genomic features and how these features are related to each other is critical to evolutionary biology. **Materials and Methods:** This study investigates the trait evolution of genomic features in *Staphylococcus*, a bacterial clade having many pathogenic species and is of medical and pharmacological interest. Data on genome size, genomic GC, number of protein-coding genes, and genomic repeat fraction for species in *Staphylococcus* genus is collected and study their trait evolution and phylogenetic corrected relationships between them with the help of whole-genome phylogenetic trees. **Results and Conclusion:** We observe that the 4 genomic features studied follow differing trait evolution models genome sizes and genomic GC showing strong phylogenetic signal supporting the early-burst model, while the number of protein-coding genes and genomic repeat fraction show phylogeny-independent trait evolution. There is a significant negative correlation between genome size and genomic GC, indicating that addition of AT-rich sequences to the genome drove the increasing genome size during the early burst of diversification in *Staphylococcus*. The lack of correlation between the genome size with genomic repeat fraction and number of coding genes indicating the sequence complexity and organismal complexity evolved independently of genome size evolution in *Staphylococci* and repeat expansion may not have contributed to the genome size increase during the diversification.

Keywords: *Staphylococcus*, Phylogenetic signal, Trait Evolution, Phylogenetic conservation.

Correspondence:
Akash Ajay
Lab 301, School of Environmental Sciences, Jawaharlal Nehru University, New Delhi-110067, INDIA.

Email: akash75_ses@jnu.ac.in

INTRODUCTION

Staphylococcus is a genus that comprises several gram-positive bacterial species of critical importance to human health and welfare.[1] *Staphylococci*, known for their adaptability, are ubiquitous in the environment and commonly found on human and animal skin and mucous membranes.[1,2] The *Staphylococcus* species are of special medical and pharmacological interest as they are noted to cause a wide variety of infections from minor skin infections to a life-threatening septal strokes.[1-3] Additionally, some strains of *Staphylococcus* have developed resistance to multiple antibiotics, posing a significant public health challenge. *Staphylococcus* is a subject of extensive microbiology and medical science research, seeking to understand its pathogenicity, antibiotic resistance mechanisms, and potential therapeutic interventions.[4]

This paper attempts to understand the trait evolution of genome features in *Staphylococcus*, which can provide crucial insights into *Staphylococcus* evolution. Genome and evolutionary processes share a bidirectional

SCAN QR CODE TO VIEW ONLINE

www.ajbls.com

DOI: 10.5530/ajbls.2023.12.69

relationship with genomic features influencing the evolutionary trajectories of the species, which in turn shape the species' genome. A species' genome can be considered as a rich record of its evolutionary history. There are several genomic features, some of the most important of which are genome size, genomic GC, genomic complexity, and genomic repeat fraction. For species with smaller genome sizes, genome size and complexity tend to be correlated.[5]

To study the influence of phylogeny on the evolution of these genomic features, I opt for the following methods commonly used in evolutionary biology, Phylogenetic Generalized Least Squares (PGLS), trait evolutionary models, and statistical tests to check for phylogenetic signals. In this paper, I will briefly describe these techniques before applying them to understand the trait evolution of genomic features in *Staphylococci*.

PHYLOGENETIC COMPARATIVE METHODS

Phylogenetic signal and trait evolution

Comparative biology frequently involves comparing the species features across taxons and inferring the trait/ morphological evolution from the variation involved in the characters. Traits could evolve around the phylogenetic tree via random walks (Brownian motion) and stasis (no real evolutionary change for long periods) or directional evolution. There are several trait evolution models,[6] of which Brownian motion and OU (Ornstein-Uhlenbeck) are the most popular ones used and found in all analyses. Special statistical methods are needed for studying comparative modelling across a lineage or examining the co-evolution of two traits, known as phylogenetic comparative methods. Felsenstein noted in his seminal 1985 paper[7] that any regression and correlation done without accounting for the phylogenetic relationship of data points was inherently flawed as the samples were not drawn independently from a random distribution as is assumed in any correlation/ regression analysis. The existence of subgroups due to the phylogenetic relationship will present a misleading picture of the traits and generate spurious correlations. Pagel[8] was one of the first to propose a statistical model- Pagel's λ to detect phylogenetic signals-a tendency of phylogenetically close species to share trait values. Pagel's λ of value 1 indicates a strong phylogenetic signal, while 0 indicates a weak signal. Another parameter of Bloomberg's constant, K, is similar to Pagels λ in its application but differs in the assumption of the underlying model. Pagels λ assumes the trait follows Brownian motion, while Bloomberg's K presupposes an underlying OU model.[9]

Phylogenetic Generalized Least Squares

Phylogenetic generalized regression was used to understand the relationship between different genomic features. Phylogenetic Generalized Least Squares (PGLS)[10] is a powerful statistical tool in comparative biology that extends the traditional Generalized Least Squares framework. PGLS incorporates the phylogenetic relatedness among species by modelling the error term's covariance structure based on the phylogenetic tree. This method is especially valuable when studying trait evolution across multiple species, as it properly accounts for the shared evolutionary history of the taxa, providing more robust and accurate parameter estimates in regression analyses. PGLS is widely used for exploring trait-environment relationships while controlling for the effects of phylogeny.

MATERIALS AND METHODS

Collection of data on genomic features of Staphylococci

Genomes of species belonging to the *Staphylococcus* genus were taken from NCBI (https://www.ncbi.nlm.nih.gov/genome/). Supplementary Table 1 lists the NCBI accession numbers of the reference genomes of the strain taken for this study. The number of coding genes was used as a proxy for genomic complexity. The genome size, genomic GC, and the number of coding genes were taken from the median count of the corresponding species. The data of the parameters is given in Supplementary Table 1.

Computation of genomic repeat fraction

The repeats in the genomes were identified with the help of the repeat finder plugin GeneiousPrime2023. The repeat finder algorithm detects the tandem repeats independent of any database and uses a k-mer-based search for repeats of a period size between 2-1000.[11] The internal repeats were excluded, and the repeat lengths were summated and divided by the total genome size to obtain the genomic fraction. The values of genomic repeat fraction are shown in Supplementary Table 1 with other features.

Plotting the Staphylococcus Phylogenetic tree

TYGS server (https://tygs.dsmz.de/) was used to construct both 16S and whole genome-based phylogenetic trees.[12] Figure 1 shows *Staphylococcus's* 16S rRNA and whole genome-based phylogenetic tree. The Newick format of the whole genome phylogenetic tree

was used for the phylogenetic comparative modelling. We have opted to use a whole-genome phylogenetic tree for phylogenetic comparative modelling in our study because of the close phylogenetic distances between the species involved

Phylogenetic comparative analysis

For statistical analysis, the programming language R v.4.01 (RC Team., 2000) was used. We used the Pearson correlation test to check for the correlations using the cor. test () function of the base package. The correlation heatmaps were drawn using ggplot2.[13] The phylogenetic signals were checked using Pagel's λ,[14] Bloomberg's K, and Abouheif test. I used the Geiger package of R for performing comparative phylogenetic modelling analysis and fit the genomic feature data on BM, OU, Early-Burst (EB), lambda and white noise models using fit Continous function. I used ape[16] and caper[17] packages in R to perform PGLS to check the cross-correlations between the genomic features. The *p*-value of PGLS was obtained by one-way ANOVA.

RESULTS

Genomic features differ in their phylogenetic signals

Using the whole-genome sequence-based phylogenetic tree, I checked the phylogenetic signal for each genomic feature. I opted not to use the 16S rRNA sequence phylogenetic tree as it had two edges of zero length, indicating that some species are too close in the evolutionary timescale; we opted to use the whole genome sequence-based tree.

The phylogenetic signal and autocorrelation for each genomic feature were checked with the help of Pagel's λ, Bloomberg's K, Moran's I, and Abouheif test, as mentioned in the methods section. The first two parameters describe the role of phylogeny in trait evolution under random walks and stabilizing selection conditions, respectively. Abouheif's test will supplement their results by describing the trait values' clustering according to the lineage's phylogeny. Table 1 gives the results of phylogenetic signal testing.

Table 1 shows the phylogenetic signals for the 4 genomic features studied. We note significant values for genome size and genomic GC Pagel's λ and Bloomberg's K implying a strong phylogenetic signal under both Brownian and OU conditions. Pagel's λ for genome size and genomic GC is nearly equal to 1, indicating a strong phylogenetic signal. Bloomberg's K is greater than 1 for genome size and GC, meaning a stronger phylogenetic signal than expected if the traits were under Brownian motion. Bloomberg's result can be interpreted as phylogenetic conservatism as evolutionarily close species have closer genome size and GC values. However, we note that coding genes proportional to genome size for prokaryotes and unicellular eukaryotes do not show any significant phylogenetic signal. No significant phylogenetic signal was observed for genomic repeat fraction. Abouheif's test agrees with Pagel's λ and Bloomberg's K tests and shows significant phylogenetic autocorrelation, implying related species will be similar in genome size and genomic GC.

Genomic features differ in their trait evolutionary models

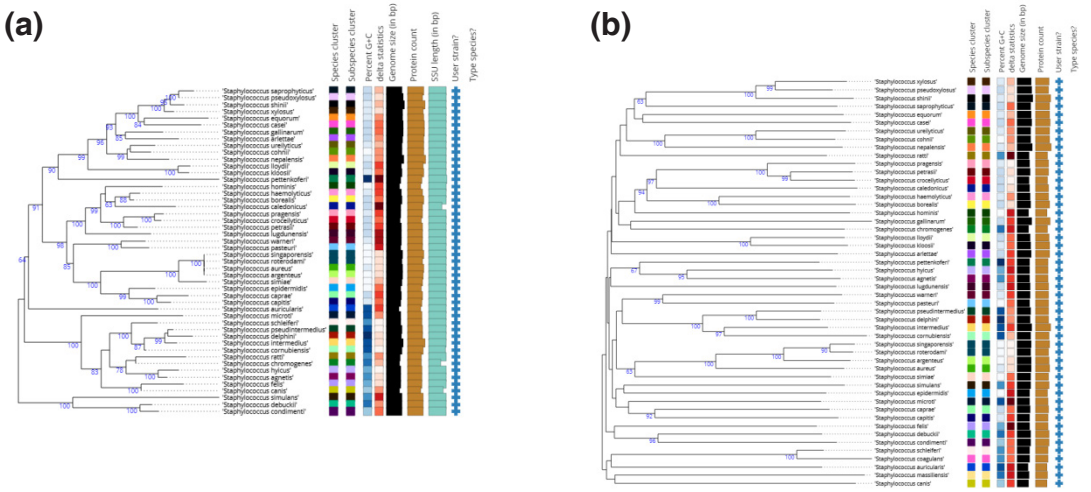

Figure 1: Phylogenetic tree of Staphylococcal species using (a) 16S rRNA sequence and (b) whole-genome sequence.

Examining the trait evolution models for each of the genomic features in *Staphylococci* bacteria, we find that while genome size and genomic GC both favour early burst models, the genomic complexity, i.e., the number of coding genes, seem to follow the lambda model, and genomic repeat fraction seems to favour either of the white noise or Brownian motion. The white noise model suggests non-directional evolution independent of phylogeny,[16] and Brownian motion[16] suggests a normal distribution of traits in the phylogeny and increasing variance as the trait evolves. Table 2 shows model fitting for trait evolution in the genomic features in *Staphylococci*. Regarding the trait evolution of genome size and genomic GC, these results are similar to Gao and Wu, 2022 who inferred the trait model of genome size and GC using a gene tree-based phylogenetic tree.

Genome size and Genomic GC are positively correlated in the Staphylococci

To compare the relationship of genomic features with each other, the Spearman correlations between the features were checked, followed by the PGLS test to check for correlations that remain after correcting for phylogeny. Figure 2 shows Spearman correlation plots for the genomic features studied in Staphylococcus with corresponding Log P plots.

I note that the genomic repeat fraction is not correlated with other genomic features, while genome size is positively correlated with coding genes and negatively correlated with genomic GC. The correlation between genome size and genomic complexity is well-known for species with smaller genomes, like prokaryotes and unicellular eukaryotes.[16] Since the correlation between the features without accounting for phylogenetic corrections can be misleading, we rechecked the correlations with PGLS. We have opted to use PGLS instead of Phylogenetic Independent Contrasts (PIC) as we note most of the genomic features do not follow the Brownian motion of trait evolution so PIC may present a misleading picture. From section 3.1, we can see that except for genomic repeat fraction, no other genomic features following Brownian motion PIC (phylogenetic independent contrast) will not be applicable.

It can be seen in Table 3, post phylogenetic corrections, that all the previously observed correlations between genomic features have become insignificant, except for the nearly significant negative correlation of genomic GC with genome size, implying the genome size burst may have been driven by the gain of AT-rich sequences. The previously observed correlation between genome size and protein-coding genes appears due to phylogeny. The insignificant *p* value between genome size and genomic repeat fractions shows that sequence complexity is not related to genomic size in *Staphylococci*, and repeat expansions are unlikely to have contributed to the evolution of genome size or complexity in *Staphylococci*.

DISCUSSION

The investigation of trait evolution in the *Staphylococcus* genus shows that both genomic repeat fraction and genomic complexity (protein-coding genes) don't show any significant phylogenetic signal while genome size and genomic GC have significant phylogenetic signals and follow early burst models as was noted by Gao *et al.* 2022[19] using gene tree based phylogenetic trees. Although model fitting shows two models-the BM model and the white model to be equally likely for genomic repeat fraction, the white model is likely true as the genomic repeat fraction showed no significant phylogenetic signal in the phylogenetic signal test. This implies that genomic repeat fraction is evolving independently of the phylogeny in *Staphylococcus*, a similar result was also noted in our other study on *Streptococcus*.[20] The number of protein-coding genes follows the lambda model in which the lambda value is zero, which implies no phylogenetic influence.[16] So, I conclude that the trait evolution of genomic features repeat fraction and a number of protein-coding genes are independent of phylogeny in *Staphylococcus*. The strong correlation between genome size and genomic complexity generally noted for microbes[20] (i., the number of protein-coding

Table 1: Phylogenetic signal detection using Pagel's λ and Bloomberg's K and Abouheif's test for genomic features in Staphylococcus.

Parameter	Pagel's λ		Bloomberg's K		Abouheif's Test	
λ		*p* value	Lambda	*p* value	Coefficient	*p* value
Genome size	0.999	1.41E-05	1.341	0.001	4.079	0.001
Genomic GC	0.999	9.78E-07	1.366	0.001	4.205	0.001
Coding genes	6.61E-05	1	0.561	0.77	-0.686	0.777
Genomic repeat fraction	0.5941	1	0.792	0.294	-0.081	0.485

Table 2: Evaluating trait models for each of the genomic features.

Trait	Model	AIC
Genome size	Brownian Motion	-41.51
	OU model	-41.51
	EB model	**-44.002**
	lambda model	-39.51
	white noise	-21.96
Genomic GC	Brownian Motion	193.606
	OU model	193.601
	EB model	**177.117**
	lambda model	195.606
	white noise	216.173
Coding genes	Brownian Motion	739.597
	OU model	755.171
	EB model	741.597
	lambda model	**739.506**
	white noise	739.597
Genomic repeat fraction	**Brownian Motion**	**357.518**
	OU model	359.925
	EB model	359.518
	lambda model	359.518
	white noise	**357.518**

genes) disappears for *Staphylococci* when corrected for phylogeny, indicating that a burst in genome size did not drive any similar increase in the genomic complexity in the genus. So, despite being correlated,[13] genome size and the number of protein-coding genes follow essentially different trait evolutionary models in the *Staphylococcus* genus .So *Staphylococcus* does not seem to follow genome size expansion or loss due to gain or loss of functions of genes as proposed by Bobay and Ochman (2007).[21] Not all bacterial genera show these trends as we noted in our previous study on *Streptococcus,* where the correlations between genome size and coding genes remain significant post-phylogenetic corrections.[20]

These kinds of results can give useful insights into the evolutionary processes that might have shaped the genome size evolution in the clade.[22] A nearly significant correlation between genome size and GC, and both follow the early burst model, both show significant phylogenetic signals indicating that *Staphylococcus*, in its early diversification stage, might have incorporated more GC-poor regions in its genome while expanding its genome. Observing an early burst model of genomic size and GC is consistent with other studies.[19,22,23] Gao *et al.*, 2022 observed the predominance of the early burst model for both genome size and genomic GC while studying the genome size and GC evolution in a broader range of microbial species using phylogenetic trees drawn from a smaller section of genes, I observed a similar result when I use the phylogenetic trees based on whole genomes The lack of correlations seen between

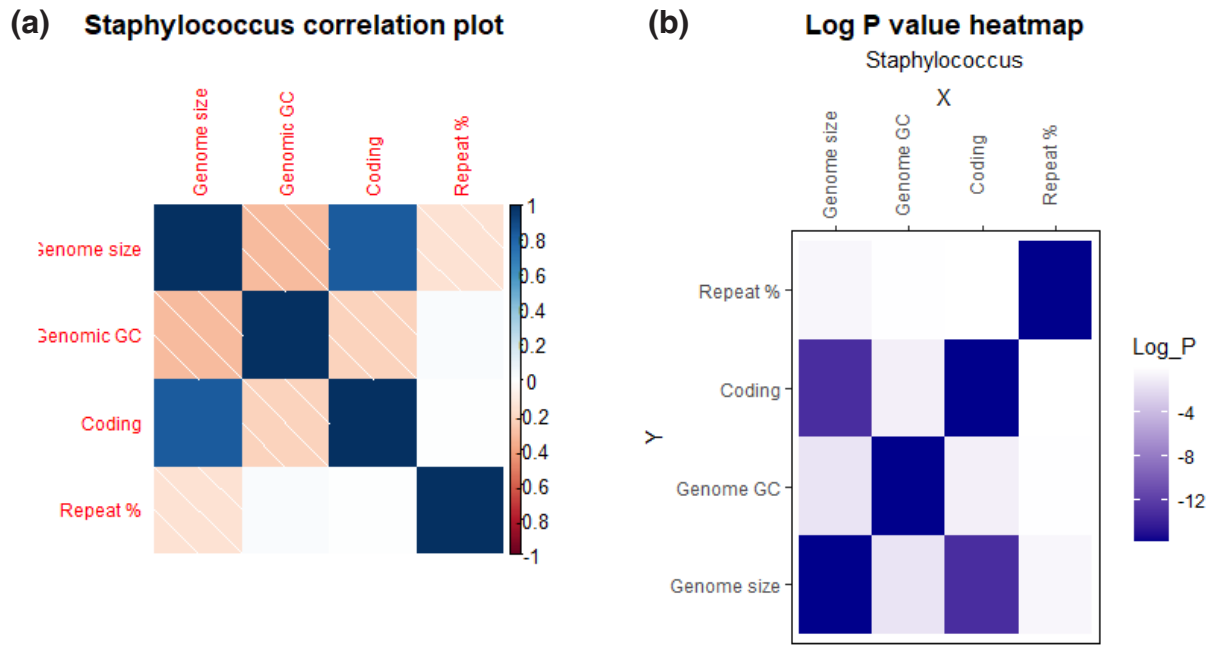

Figure 2: Correlation plot for genomic features-genome size, genomic GC, number of protein-coding genes, and genomic repeat fraction. The Log *p* value heat map corresponds to the correlation plot/heat map showing the significance of p values.

Table 3: PGLS tests to see the relationship of genomic features with each other.

Parameters compared	slope	*p*-value	AIC	RMS value	Kappa	Lambda	Delta
Size vs. GC	-2.702	0.071	188.179	25.541	1	1	1
Size vs. coding	-9.73	0.653	357.754	758.843	1	1	1
Size vs. Repeat fraction	0.891	0.228	357.869	760.596	1	1	1
GC vs. coding	-2.702	0.852	188.179	25.541	1	1	1
GC vs. Repeat fraction	-9.730	0.247	357.754	758.843	1	1	1
Repeat vs. coding	0.891	0.667	357.869	760.596	1	1	1

genome size and genomic repeat size may also imply the repeat expansion or low complexity sequences may not have played any role in the genome size evolution This is unlike mammals and mammals where the genomic repeats have been one of the factors supporting the genome size evolution driving its increase.[24]

It can also guide similar studies to solve similar questions regarding the genomics, niche, and lifestyles of the pathogenetic bacterium by applying a phylogenetic/evolutionary approach to the problem. One limitation of this study is that non-genomic factors like population size[25] and environmental factors like thermal stress[26] are also known to influence the genome size evolution in bacteria. They may contribute to the evolution of genome size in Staphylococcus, which can be explored in further studies. Another limitation of this study is that this study limits itself to examining the trends across the species and does not include information at the strain level as we cannot compute whole-genome-based phylogeny trees for more than 50 species at a time on the TYGS server. This study can help understand the macro-evolution of genomic traits in *Staphylococcus* and be a helpful guide to more such studies on pathogenic and commensal bacteria which are important for human health and disease. It can also guide similar studies to solve similar questions regarding the genomics, niche, and lifestyles of the pathogenetic bacterium by applying a phylogenetic/evolutionary approach to the problem.

ACKNOWLEDGEMENT

I thank Professor Shandar Ahmad and Professor Krishan Kumar of Jawaharlal Nehru University for their discussions and suggestions. I thank Dr. Tina Begum, Swiss Institute of Bioinformatics for discussions on PGLS interpretations. I thank UGC for the SRF fellowship.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

FUNDING

No funding source was received for this study

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable

PATIENT CONSENT

Not applicable

ABBREVIATIONS

BM model: Brownian motion model; **OU model:** Ornstein Uhlenbeck model; **EB model:** Early burst model; **PGLS:** Phylogenetic generalized least squares; **AIC:** Akaike Information criterion; **RMS:** Root mean square; **ANOVA:** Analysis of Variance; **PIC:** Phylogenetic independent contrasts.

SUMMARY

The genomic, morphological, and developmental features of species are constrained by the phylogeny and are of central interest in ecological, evolutionary, and pharmacological research. This study takes 4 global genomic features: genome size, genomic GC, number of protein-coding sequences and genomic repeat fraction and studies their evolution in the *Staphylococcus* genus, a bacterial clade having bacterial species of medical and pharmacological research. In this study we explore the following questions: Do these 4 genomic features follow similar models of trait evolution? Are these features related to each other, given their phylogenetic non -independence? Do the results obtained with the help of whole-genome-based: phylogenetic trees agree with the results inferred from phylogenetic trees made by fewer selections of genes? The information on genome size, GC, and the number of protein-coding genes came from the NCBI website, from where the genomes were downloaded, while the genomic repeat fraction was computed with the help of the repeat-

finder of GeneiousPrime 2023 software. Fitting these features onto a whole-genome phylogenetic tree and examining the performance of trait evolution models: BM, OU, EB, white noise, and lambda post-checking for phylogenetic signals using Pagels λ, Bloomberg's K, and Moran's I, it was observed that those genomic features are not similar to each other in their trait evolutionary models: while genome size and GC follow an early burst model, as observed for some microbial clades, we find that the number of coding sequences and genomic repeat fraction evolved independent of phylogeny. The number of protein-coding sequences and genome size which have been previously noted to be related for prokaryotes and lower unicellular eukaryotes, are found to be unrelated in Staphylococcus post phylogenetic correction. I also find a significant negative relationship between genome size and genomic GC which may indicate that increases in genome sizes may have been driven by the gain of AT-rich sequences rather than repeat expansion across the *Staphylococci*. These results can help in gaining insights into the genome evolution in *Staphylococci*.

REFERENCES

1. Lowy FD. Staphylococcus aureus infections. N Engl J Med. 1998;339(8):520-32. doi: 10.1056/NEJM199808203390806, PMID 9709046.
2. Otto M. Staphylococcus epidermidis-the accidental pathogen. Nat Rev Microbiol. 2009;7(8):555-67. doi: 10.1038/nrmicro2182, PMID 19609257.
3. Tang YW, Stratton CW. Staphylococcus aureus: an old pathogen with new weapons. Clin Lab Med. 2010;30(1):179-208. doi: 10.1016/j.cll.2010.01.005, PMID 20513547.
4. Novick RP, Geisinger E. Quorum sensing in staphylococci. Annu Rev Genet. 2008;42:541-64. doi: 10.1146/annurev.genet.42.110807.091640, PMID 18713030.
5. Boussau B, Daubin V. Genomes as documents of evolutionary history. Trends Ecol Evol. 2010;25(4):224-32. doi: 10.1016/j.tree.2009.09.007, PMID 19880211.
6. Münkemüller T, Lavergne S, Bzeznik B, Dray S, Jombart T, Schiffers K, *et al*. How to measure and test phylogenetic signal. Methods Ecol Evol. 2012;3(4):743-56. doi: 10.1111/j.2041-210X.2012.00196.x.
7. Felsenstein J. Phylogenies and the comparative method. Am Nat. 1985;125(1):1-15. doi: 10.1086/284325.
8. Pagel M. Inferring the historical patterns of biological evolution. Nature. 1999;401(6756):877-84. doi: 10.1038/44766, PMID 10553904.
9. Blomberg SP, Rathnayake SI, Moreau CM. Beyond Brownian motion and the Ornstein-Uhlenbeck process: stochastic diffusion models for the evolution of quantitative characters. Am Nat. 2020;195(2):145-65. doi: 10.1086/706339, PMID 32017624.
10. Mundry R. Statistical issues and assumptions of phylogenetic generalized least squares. Mod phylogenetic Comp Methods Appl Evol Biol Concepts Pract. 2014:131-53.
11. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573-80. doi: 10.1093/nar/27.2.573, PMID 9862982.
12. Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun. 2019;10(1):2182. doi: 10.1038/s41467-019-10210-3, PMID 31097708.
13. Wickham H. ggplot2. WIREs Computational Stats. 2011;3(2):180-5. doi: 10.1002/wics.147.
14. Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, *et al*. geiger v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics. 2014;30(15):2216-8. doi: 10.1093/bioinformatics/btu181, PMID 24728855.
15. Mundry R. Statistical issues and assumptions of phylogenetic generalized least squares. Mod phylogenetic Comp Methods Appl Evol Biol Concepts Pract. 2014:131-53.
16. Gregory TR. Genome size and developmental complexity. Genetica. 2002;115(1):131-46. doi: 10.1023/a:1016032400147, PMID 12188045.
17. Paradis E, Blomberg S, Bolker B, Brown J, Claude J, Cuong HS, *et al*. Package 'ape'. Analyses of phylogenetics and evolution, version. 2019 Mar 13;2(4):47.
18. Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N, *et al*. The caper package: comparative analysis of phylogenetics and evolution in R. R package version. 2013;5(2):1-36.
19. Gao Y, Wu M. Microbial genomic trait evolution is dominated by frequent and rare pulsed evolution. Sci Adv. 2022;8(28):eabn1916. doi: 10.1126/sciadv.abn1916, PMID 35857501.
20. Ajay A. Phylogenetic comparative methods demonstrate the non-directional evolution of genomic features in the streptococcus genus. J Adv Zool. 2023;44(S-5):240-5. doi: 10.17762/jaz.v44iS-5.714.
21. Bobay LM, Ochman H. The evolution of bacterial genome architecture. Front Genet. 2017;8:72. doi: 10.3389/fgene.2017.00072, PMID 28611826.
22. Nishida H. Evolution of genome base composition and genome size in bacteria. Front Microbiol. 2012;3:420. doi: 10.3389/fmicb.2012.00420, PMID 23230432.
23. Blommaert J. Genome size evolution: towards new model systems for old questions. Proc R Soc Lond B;2020;287(1933):20201441.
24. Kapusta A, Suh A, Feschotte C. Dynamics of genome size evolution in birds and mammals. Proc Natl Acad Sci U S A. 2017;114(8):E1460-9. 25. doi: 10.1073/pnas.1616702114, PMID 28179571. Batut B, Knibbe C, Marais G, Daubin V. Reductive genome evolution at both ends of the bacterial population size spectrum. Nat Rev Microbiol. 2014;12(12):841-50. doi: 10.1038/nrmicro3331, PMID 25220308.
25. Sabath N, Ferrada E, Barve A, Wagner A. Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. Genome Biol Evol. 2013;5(5):966-77. doi: 10.1093/gbe/evt050, PMID 23563968.

Cite this article: Ajay A. Phylogeny Influences Genome Size and GC but not Sequence and Organismal Complexity in *Staphylococci*. Asian J Biol Life Sci. 2023;12(3):523-9.

Supplementary Table

Species	Accession number	Genome size (MB)	Genome GC	Number of coding genes	genomic repeat fraction
Staphylococcus agnetis	ASM1146685v1	2.474	35.7	2332	10.307
Staphylococcus argenteus	ASM23692v1	2.784	32.3	2542	15.014
Staphylococcus arlettae	41556_H01	2.601	33.468	2453	20.261
Staphylococcus aureus	ASM1342v1	2.8278	32.7	2674	14.888
Staphylococcus auricularis	ASM1602829v1	2.251	37.1	2046	10.573
Staphylococcus borealis	7067_4#69	2.629	33.7	2350	11.754
Staphylococcus caledonicus	ASM1623846v1	2.503	33.6	2339	8.75
Staphylococcus canis	ASM1623844v1	2.229	34.8	2043	15.747
Staphylococcus capitis	ASM2527281v1	2.506	32.8	2334	10.455
Staphylococcus caprae	ASM396662v1	2.629	33.5	2428	13.96
Staphylococcus casei	ASM3029440v1	2.913	33.26	2693	15.517
Staphylococcus chromogenes	ASM299430v1	2.323	36.6	2198	7.792
Staphylococcus coagulans	ASM2255713v1	2.471	35.9	2238	4.411
Staphylococcus cohnii	44343_D01	2.643	32.337	2447	20.621
Staphylococcus condimenti	ASM192240v1	2.685	34.6	2486	13.296
Staphylococcus cornubiensis	SAMEA104055222	2.677	37.4	2412	10.347
Staphylococcus croceilyticus	ASM468487v1	2.378	33.3	2218	9.672
Staphylococcus debuckii	ASM371873v1	2.691	36.6	2480	13.043
Staphylococcus delphini	ASM2555882v1	2.579	38.1	2370	14.541
Staphylococcus epidermidis	ASM609437v1	2.505	32	2267	19.042
Staphylococcus equorum	ASM1612745v1	2.765	33	2612	12.297
Staphylococcus felis	ASM301291v1	2.395	34.9	2240	38.747
Staphylococcus gallinarum	ASM2079015v1	2.925	33.1	2715	7.419
Staphylococcus haemolyticus	ASM161195v1	2.492	32.7	2360	23.034
Staphylococcus hominis	ASM381250v1	2.25	31.4	2138	20.889
Staphylococcus hyicus	ASM81608v1	2.492	35.7	2336	11.477
Staphylococcus intermedius	42197_D02	2.68914	37.45	2498	29.675
Staphylococcus kloosii	ASM301925v1	2.636	32.85	2522	14.226
Staphylococcus lloydii	ASM1577597v1	2.544	33.3	2409	12.854
Staphylococcus lugdunensis	ASM155877v1	2.5717	33.7	2367	14.66
Staphylococcus massiliensis	ASM29807v1	2.366	36.5	2148	10.989
Staphylococcus microti	50432_E01	2.409	38	2287	54.047
Staphylococcus nepalensis	ASM244293v1	2.886	33	2678	17.394
Staphylococcus pasteuri	ASM1659979v1	2.543	31.5	2387	10.106

Supplementary Table

Species	Accession number	Genome size (MB)	Genome GC	Number of coding genes	genomic repeat fraction
Staphylococcus petrasii	50305_H01	2.575	33.3	2418	18.408
Staphylococcus pettenkoferi	ASM220880v2	2.4723	38.8	2271	15.775
Staphylococcus pragensis	ASM478566v1	2.435	33.1	2312	8.501
Staphylococcus pseudintermedius	ASM1612671v1	2.641	37.4	2425	15.752
Staphylococcus pseudoxylosus	ASM1850196v1	2.925	34.088	2586	8.615
Staphylococcus ratti	ASM2088353v1	2.323	36.085	2153	16.832
Staphylococcus roterodami	EMCR19.fasta.gz	2.705	32.3	2434	9.649
Staphylococcus saprophyticus	ASM781411v1	2.635	33	2455	13.359
Staphylococcus schleiferi	MGYG-HGUT-01437	2.435	35.9	2117	19.425
Staphylococcus shinii	ASM1758306v1	3.007	32.5	3.007	6.186
Staphylococcus simiae	50377_D01	2.555	31.9	2211	23.875
Staphylococcus simulans	ASM2556146v1	2.66	36	2472	14.474
Staphylococcus singaporensis	ASM1502507v1	2.735	32.2	2428	8.154
Staphylococcus ureilyticus	ASM2555884v1	2.692	32.5	2597	13.373
Staphylococcus warneri	ASM357172v1	2.54	32.6	2402	15.394
Staphylococcus xylosus	ASM70941v1	2.816	32.7	2599	11.541