

Codon Usage Bias and Evolutionary Determinants of Codon Preference in *Campylobacter jejuni*

Samira Alam¹, Syed Hasib Akhter Faruqui², and Anand B. Karki¹

¹ Department of Biology, Sam Houston State University, Tx, USA

² Department of Engineering Technology, Sam Houston State University, Tx, USA

1. Abstract

Campylobacter jejuni is a major foodborne bacterial pathogen whose genome exhibits substantial diversity between core and accessory gene sets. This study aimed to characterize codon usage patterns in core and accessory genes and evaluate the genomic and evolutionary factors influencing codon preference, including guanine-cytosine (GC) composition, mutational pressure, and natural selection. Out of the 114,617 *C. jejuni* genomes available in the NCBI database; genomes with $\geq 99.9\%$ nucleotide identity were considered near-duplicates and removed to minimize redundancy. A total of **305 non-redundant genomes** were retained for subsequent pangenome and codon-usage analyses. Genomes were annotated using Bakta. Comparative pangenome analysis was performed to classify genes into core, shell, and cloud categories. Codon usage was evaluated using relative synonymous codon usage (RSCU), effective number of codons (ENC), codon adaptation index (CAI), tRNA adaptation index (tAI), and GC content. RSCU analysis revealed marked preferences among synonymous codons, with AGA being highly preferred and CGG being infrequently used for arginine. A mean CAI of 0.67 and a mean ENC of approximately 39 indicated moderate codon usage bias, whereas the relatively low mean tAI of 0.36 suggested weak adaptation to the cellular tRNA pool. CAI was inversely associated with GC content, and most genes clustered around an ENC of approximately 39 at relatively low GC3 values, supporting an association between nucleotide composition and codon usage bias. Furthermore, neutrality plot analysis suggested a stronger influence of natural selection on codon usage bias than mutational pressure. Overall, these findings indicate the combined effects of nucleotide composition and evolutionary effect, with natural selection being predominant force shapes the codon preference of *C. jejuni*.

2. Objectives

- To characterize codon usage bias (CUB) in *Campylobacter jejuni*
- To identify the evolutionary and genomic factors that influence codon preference

3. Methods

- 305 complete genomes analyzed
- Downloaded from NCBI database using FTP Server
- Genes containing ≥ 100 codons included

3.1 Pangenome Analysis Workflow

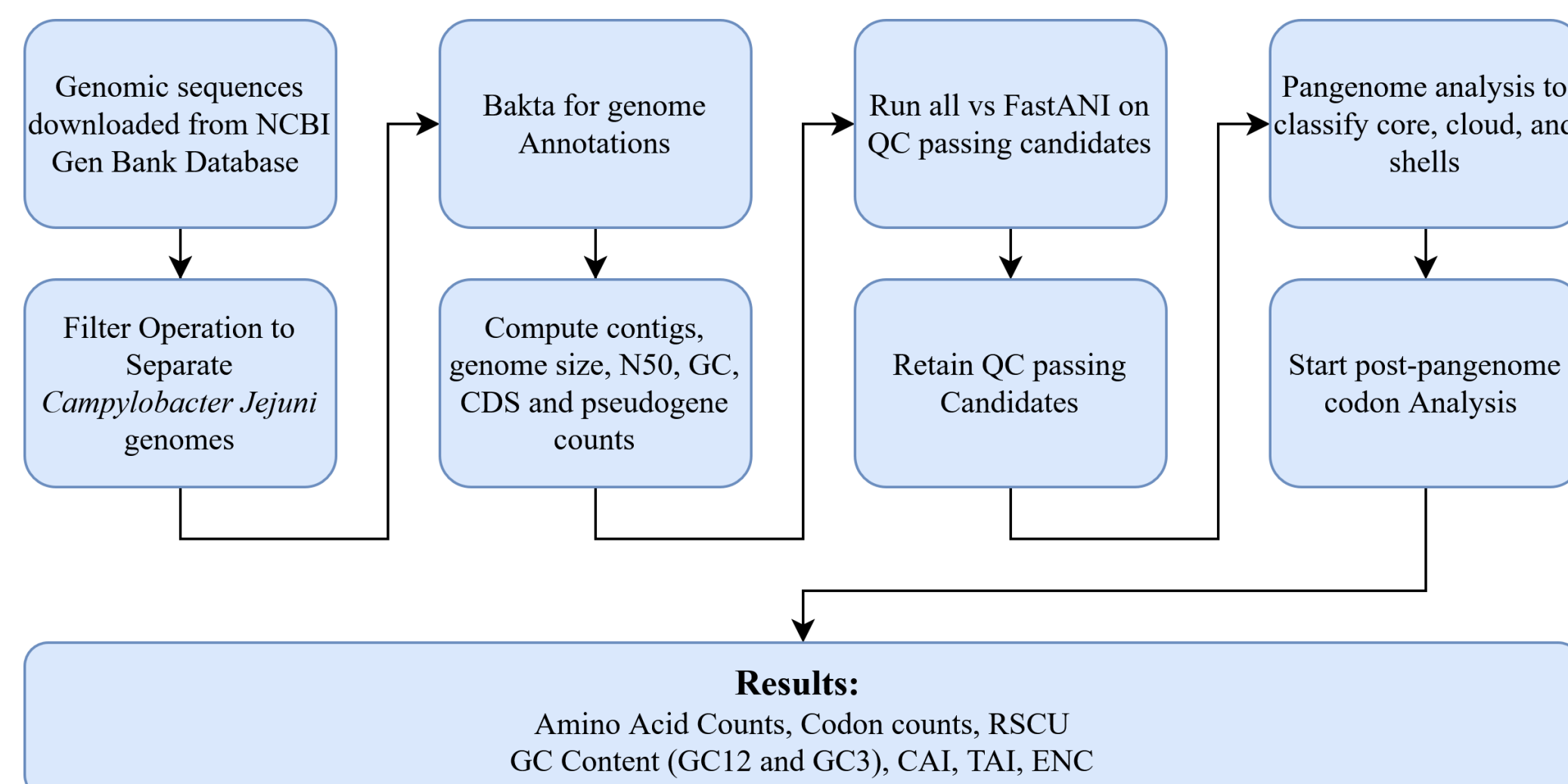


Figure 1:Pangenome Analysis Workflow

3.2 *Campylobacter jejuni* Gene Partition

- Pan-genome continues to increase as additional *C. jejuni* genomes are included.
- Core genome decreases as more *C. jejuni* genomes are added, then the level stabilizes.

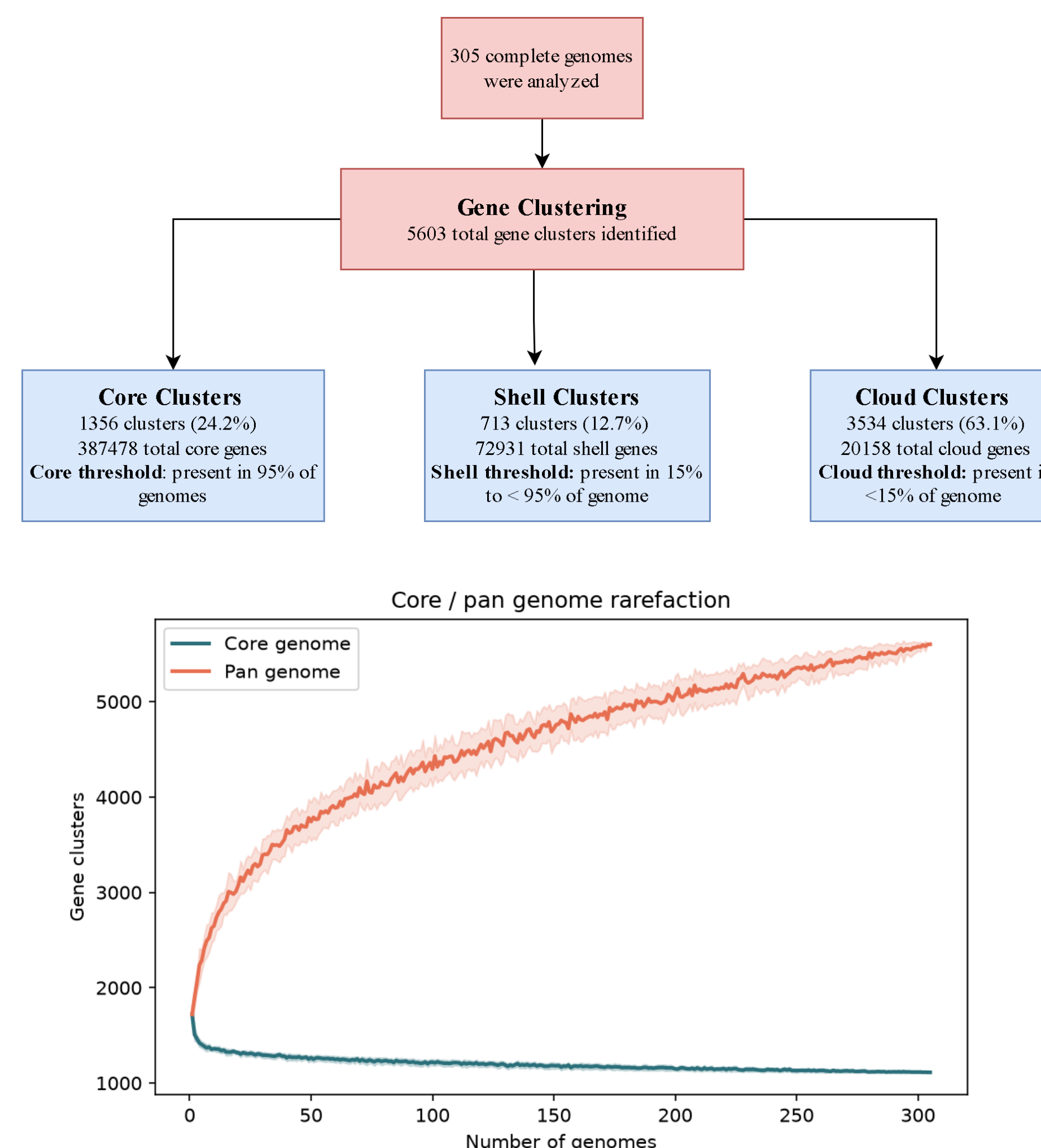


Figure 2:Core/ pan genome rarefaction

3.3 Codon Usage Bias with RSCU values

Amino Acid	Codon Core Genes	Count	RSCU
Ala	GCA	3018810	1.317336
Ala	GCC	866459	0.378102
Ala	GCG	727317	0.317384
Ala	GCT	4553824	1.987179
Cys	TGC	464317	0.583351
Cys	TGT	1118647	1.413357
Asp	GAC	536573	0.157018
Asp	GAT	6297989	1.842982
Glu	GAA	7593688	1.639213
Glu	GAG	1671351	0.360787
Phe	TTC	563810	0.144151
Phe	TTT	7258706	1.855849
Gly	GGA	2771456	1.491336
Gly	GGC	1125515	0.605646
Gly	GGG	772157	0.415502
Gly	GGT	2764355	1.487515

For Arginine, AGA shows highest codon preference and **CGG** shows lowest codon preference in both core and accessory genes

Core Genes

Top 5 favored Codons with RSCU	Top 5 Avoided Codons with RSCU
AGA (Arg) : 3.1488	CGG (Arg): 0.0201
TTA (Leu) : 2.7591	CTG (Leu): 0.0537
CCT (Pro) : 2.3414	TCG (Ser): 0.1414
AGT (Ser) : 2.1366	TTC (Phe): 0.1444
GCT (Ala) : 1.9884	CTC (Leu): 0.1460

Accessory Genes

Top 5 favored Codons with RSCU	Top 5 Avoided Codons with RSCU
AGA (Arg): 3.7315	CGG (Arg): 0.0513
TTA (Leu): 2.8929	CTG (Leu): 0.0748
CCT (Pro): 2.2435	TCG (Ser): 0.1481
GCT (Ala): 2.8929	TTC (Phe): 0.1584
AGT (Ser): 1.9477	CTC (Leu): 0.1675

3.4 Mean CAI, ENC and tAI across Core, Shell and Cloud

- CAI** (≈ 0.67): Moderate codon usage bias; preferred codons coexist with less-preferred synonymous codons, lowering the overall CAI.
- ENC** (≈ 39): Moderate codon usage bias, indicating genome favors certain synonymous codons, but not to the extreme.
- tAI** (≈ 0.36): Relatively low tRNA adaptation, suggesting codon usage is not strongly matched to the available tRNA pool.

Partition	ENC mean difference	P-value for ENC	CAI mean difference	P-value for CAI	tAI mean difference	P-value for tAI
core vs cloud	.5636	0.0000	.0102	0.0000	9.8×10^{-5}	0.6626
shell vs cloud	.6201	0.0000	.0013	0.0007	7.3×10^{-4}	0.0000
shell vs core	.0565	0.0004	.0115	0.0000	8.3×10^{-4}	0.0000

3.5 Relationship among CAI, GC3 and ENC

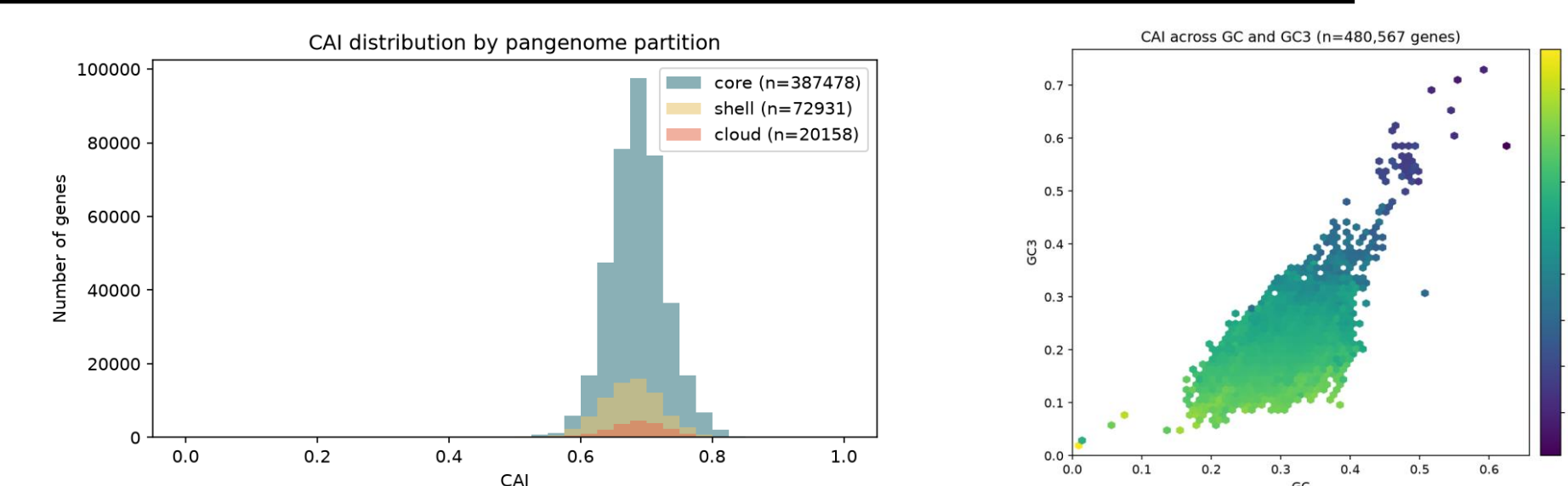


Figure 3: CAI increases as GC3 decreases, showing an inverse relationship between codon adaptation and nucleotide composition

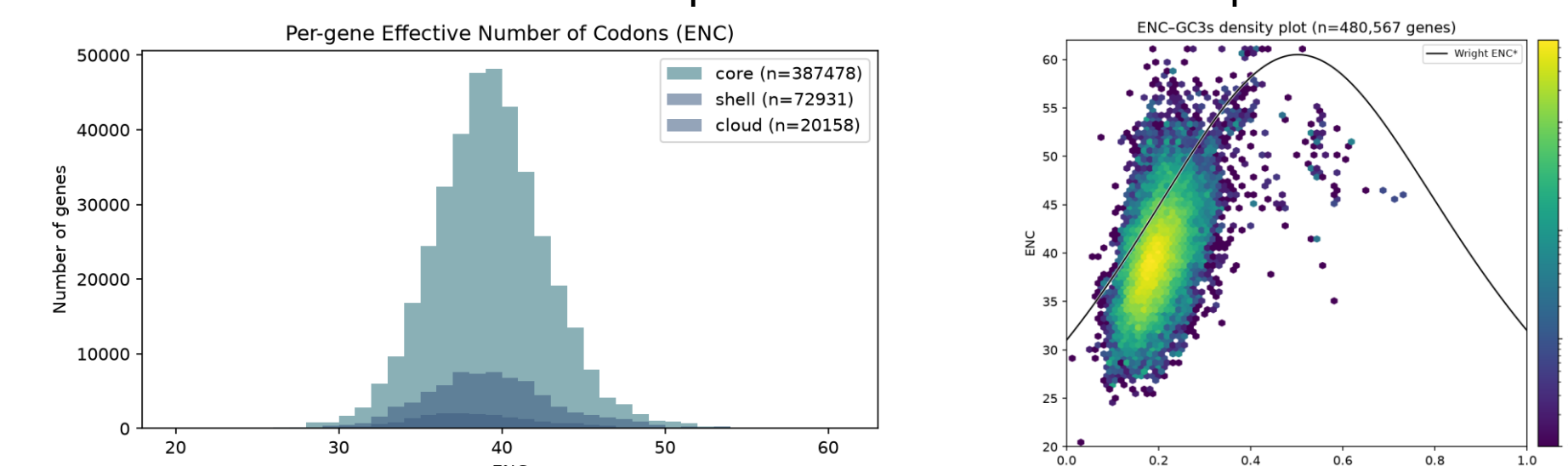
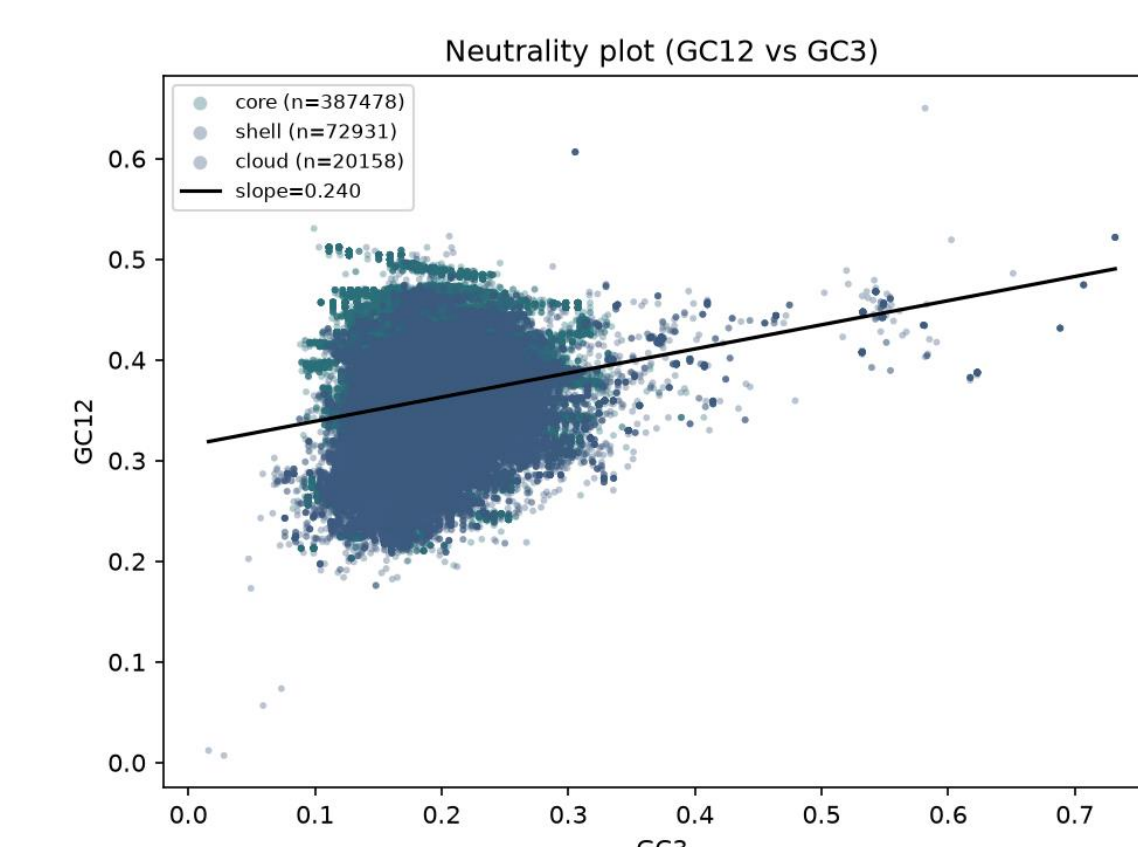


Figure 4: Most genes cluster around ENC 35–45 and low GC3, suggesting a relationship between nucleotide composition and codon usage bias

3.6 Neutrality Plot

- A low regression slope 0.198 indicates weak GC12–GC3 correlation, suggesting that natural selection has a stronger influence than mutation pressure on codon usage across all gene categories.



4. Discussion

The codon usage analysis indicates moderate codon usage bias in *C. jejuni*, as reflected by the CAI (~ 0.67) and ENC (~ 39) values. Clustering of genes around ENC 35–45 and low GC3 supports an association between nucleotide composition and codon usage patterns. The inverse relationship between CAI and GC3 suggests that AT-rich genes tend to exhibit greater codon adaptation, indicating that nucleotide composition contributes to variation in codon usage patterns. The low GC12–GC3 regression slope suggests a weak relationship between these compositional measures, indicating that factors other than mutational pressure, including natural selection, may contribute substantially to codon usage variation across the analyzed genes.