

Large-scale analysis of SARS-CoV-2 synonymous mutations reveals the adaptation to the human codon usage during the virus evolution

Daniele Ramazzotti,^{1,*,†,§} Fabrizio Angaroni,^{2,†} Davide Maspero,^{2,3} Mario Mauri,¹ Deborah D'Aliberti,¹ Diletta Fontana,^{1,*} Marco Antoniotti,^{2,4} Elena Maria Elli,^{1,5,†} Alex Graudenzi,^{3,4,†} and Rocco Piazza^{1,4,*}

¹Department of Medicine and Surgery, Università degli Studi di Milano-Bicocca, Via Cadore 48, Monza 20900, Italy, ²Department of Informatics, Systems and Communication, Università degli Studi di Milano-Bicocca, Viale Sarca 336, Milano 20100, Italy, ³Institution of Molecular Bioimaging and Physiology, Consiglio Nazionale delle Ricerche (IBFM-CNR), Via fratelli Cervi 93, Segrate 20054, Italy, ⁴Bicocca Bioinformatics, Biostatistics and Bioimaging Center—B4, Veduggio al Lambro, Via Follereau 3, 20854, Italy and ⁵Hematology and Clinical Research Unit, San Gerardo Hospital, Via G. B. Pergolesi, 33, Monza 20900, Italy

†Co-senior authors.

‡Equally contributing authors.

§<https://orcid.org/0000-0002-6087-2666>

*<https://orcid.org/0000-0003-1633-6310>

*Corresponding authors: E-mail: daniele.ramazzotti@unimib.it; rocco.piazza@unimib.it

Abstract

Many large national and transnational studies have been dedicated to the analysis of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) genome, most of which focused on missense and nonsense mutations. However, approximately 30 per cent of the SARS-CoV-2 variants are synonymous, therefore changing the target codon without affecting the corresponding protein sequence. By performing a large-scale analysis of sequencing data generated from almost 400,000 SARS-CoV-2 samples, we show that silent mutations increasing the similarity of viral codons to the human ones tend to fixate in the viral genome overtime. This indicates that SARS-CoV-2 codon usage is adapting to the human host, likely improving its effectiveness in using the human aminoacyl-tRNA set through the accumulation of deceitfully neutral silent mutations.

One-Sentence Summary. Synonymous SARS-CoV-2 mutations related to the activity of different mutational processes may positively impact viral evolution by increasing its adaptation to the human codon usage.

Key words: viral adaptation; intra-host variants; codon usage; SARS-CoV-2; viral genomics.

The COVID-19 pandemic caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has currently affected approximately 433 million people, with at least 5.9 million COVID-related deaths (WHO 2022; updated on 8 March 2022), hence posing a major health threat on a global scale.

This led to a surge of active research aimed at studying SARS-CoV-2 biology: many large national and transnational efforts were dedicated to the analysis of SARS-CoV-2 genomes, with more than 1 million sequencing datasets currently available. These analyses succeeded in reconstructing the phylogeny of the virus and, at least in part, implementing effective lineage tracking strategies (Forster et al. 2020; Seemann et al. 2020; Zehender et al. 2020; Graudenzi et al. 2021; Lemieux et al. 2021; Ramazzotti et al. 2021). Notably, only a relatively small subset of the nonsynonymous mutations detected so far, with a few relevant exceptions such as the D614G Spike mutation (Korber et al. 2020), and the mutations associated with lineages B.1.1.7,

B.1.1.28, P1, B.1.617.2, and B.1.1.529 (CDC 2021; Grubaugh et al. 2021; Muik et al. 2021; Williams and Burgers 2021) to name the major ones, were unequivocally shown to play a functional role. This evidence led to the conclusion that the majority of the reported SARS-CoV-2 sequence changes are neutral or detrimental, therefore having limited or no impact on the pathogenicity of the Coronavirus Disease 2019 pandemic (Wang et al. 2021).

To the best of our knowledge, most of the studies performed so far have focused on missense and nonsense mutations, i.e. mutations that either cause a change in the sequence of the target protein or generate a premature termination codon. However, a large fraction of the SARS-CoV-2 mutations (approximately ~30 per cent in our datasets) is synonymous, therefore changing the target codon without affecting the associated protein sequence on mRNA translation, as opposed to missense/nonsense variants.

This bulk set of mutations is usually discarded during variant analyses, as they are considered functionally irrelevant. However, it is known from decades (Ikemura 1981; Ikemura 1982; Sharp, Tuohy, and Mosurski 1986; Yang and Nielsen 1998; McCarthy, Carrea, and Diambra 2017; Dhindsa et al. 2020) that silent mutations may have a profound impact on protein expression by, i.e. changing the aminoacyl-tRNA molecule responsible for translating the mutated codon into the corresponding amino acid, a well-known process coordinated by the ribosomal machinery. A silent substitution causing a switch to a more abundant aminoacyl-tRNA is prone to increase the overall protein translating efficiency, while the opposite may occur if the tRNA pairing with the mutated codon is less abundant than the original one (Alonso and Diambra 2020; Duret 2020; Kanaya et al. 2001; Comeron 2004; Lavner and Kotlar 2005). Therefore, we hypothesize that synonymous SARS-CoV-2 variants may play a functional role by adapting the sequence of the virus to the codon usage (CU) of the human host, thus improving its effectiveness in using the human aminoacyl-tRNA set, ultimately making the translation of viral particles more efficient (Chen et al. 2020).

Notice that SARS-CoV-2 variants might be classified into *minor variants*, i.e. detected with a low variant frequency (VF) within a host (i.e. VF < 90 per cent) and *fixed variants*, i.e. detected with high VF (i.e. VF > 90 per cent). VF profiles can be derived via variant calling from raw sequencing data of viral samples, as proposed, in (Graudenzi et al. 2021; Ramazzotti et al. 2021).

Owing to the complex viral transmission dynamics involving bottlenecks and founder effects (Graudenzi et al. 2021; Ramazzotti et al. 2021), mutations, originating in a small fraction of the virion pool, are fixed overtime only in the presence of positive selection, otherwise, they tend not to fixate in the population.

To investigate the impact of CU adaptation to the SARS-CoV-2 virus evolution, we analyzed a total of 390,899 viral samples and 3,178,178 silent viral mutations identified in almost 200 different studies (see Materials and Methods), showing that:

1. The viral adaptation to the human CU is showing an increasing trend overtime during the pandemic.
2. Mutations with high codon adaptation are preferred and have a higher chance of fixating in the (intra-host) population.

To analyze a possible trend leading to the adaptation of the viral codons to the human CU during the COVID-19 pandemic, we initially analyzed a subset of high-quality samples from North America (Dataset 1, see Materials and Methods). This dataset comprises a total of 213,737 COVID-19 cases, where viral SARS-CoV-2 RNA was sequenced using an Amplicon strategy with high coverage. After alignment of the raw viral data to the SARS-CoV-2-ANC reference genome (Ramazzotti et al. 2021) and variant calling (see Materials and Methods for further details), silent variants were annotated in terms of reference and mutated codon, VF, and sampling time (Supplementary Table S1), with 'reference codon' indicating the codon found in the SARS-CoV-2-ANC reference genome and 'mutated codon' the one generated by the presence of a single nucleotide variant. The relative codon usage (RCU) adaptation was calculated as the ratio between the human CU of the mutated codon and the human CU of the wild-type counterpart, with values > 1 indicating positive adaptation.

We analyzed the RCU profile of mutated vs wild-type codons in the North America subgroup. Interestingly, we noticed that the APOBEC variants (C > T and G > A mutations) are both consistently

associated with a negative value of log₂-transformed RCU (see Fig. 1 and Supplementary Figs S1 and S2, median value -0.30), i.e. leading to a reduction in the adaptation to the human CU. As previously shown by our team and by others (Woo et al. 2007; Stavrou and Ross 2015; Milewska et al. 2018; Graudenzi et al. 2021), a relevant subset of minor variants, linked to the C > T substitution type, are most likely generated by the activation of the human APOBEC machinery (despite other mutational processes cannot be excluded). APOBEC represents a family of cytidine deaminases playing a critical role in the intrinsic responses of the host to infections operated by a large set of viruses, among which retroviruses, herpesviruses, papillomaviruses, parvoviruses, hepatitis B virus, and retrotransposons. APOBEC cytidine deaminases target the viral genome, eventually causing the functional inactivation of the pathogen. Therefore, APOBEC mutations globally represent a defense raised by the host cells to suppress viral infection (Chen et al. 2020).

In Fig. 1, we report the log₂-transformed RCU for all the 12 possible mutation classes, i.e. all the possible mutations involving the four bases (Graudenzi et al. 2021). As one can notice, different mutation classes may constitutively show different trends in terms of RCU. As an example, let us consider the variants likely linked to mutations generated by the RNA-specific adenosine deaminase ADAR (even if other mutational processes cannot be excluded) (Graudenzi et al. 2021), which show an opposite behavior compared to the APOBEC ones. ADAR represent one of the other major observed sources of mutations in SARS-CoV-2 genomes and are associated with A > G and T > C mutations (Graudenzi et al. 2021); such variants show a positive trend of adaptation to the human CU with a median value of +0.58 log₂-transformed RCU (see Fig. 1 and Supplementary Figs S1 and S2; APOBEC vs ADAR log₂-transformed RCU t-test P-value < 0.001), indicating that synonymous mutations related to different mutational processes may lead to very different effects in terms of CU.

A similar pattern was detected in two additional datasets: the first one (Dataset 2, Supplementary Figs S1 and S2) comprising 118,386 samples from the United Kingdom and the second one (Dataset 3, Supplementary Figs S1 and S2) including 58,776 samples from the rest of the world (see Materials and Methods for details). All the analyses corroborated our hypothesis, confirming in these two additional datasets all the results and conclusions discussed above.

Furthermore, these results do not change significantly when the CoCoPUTs human CU reference dataset from the Hive Lab (<https://hive.biochemistry.gwu.edu/cuts/about>; Athey et al. 2017; Alexaki et al. 2019) is considered for the analyses in place of the classical Kazusa CU reference (<https://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=9606>; Nakamura, Gojobori, and Ikemura 2000).

As different mutational processes may lead to different effects in terms of CU, we hypothesized that an RCU less fit to the human translational machinery could cause mutated SARS-CoV-2 genomes to reduce their translation efficiency, therefore undergoing purifying selection. As opposite, the variants leading to a better match with the human CU should undergo positive selection, overcome translational bottlenecks and outcompete wild-type viruses in protein synthesis, and ultimately improve the efficiency of viral packaging. Although a direct demonstration of this hypothesis is challenging, a greater RCU should correspond to increased viral fitness, therefore potentially translating into a greater VF at the global level. To shed light on this phenomenon,

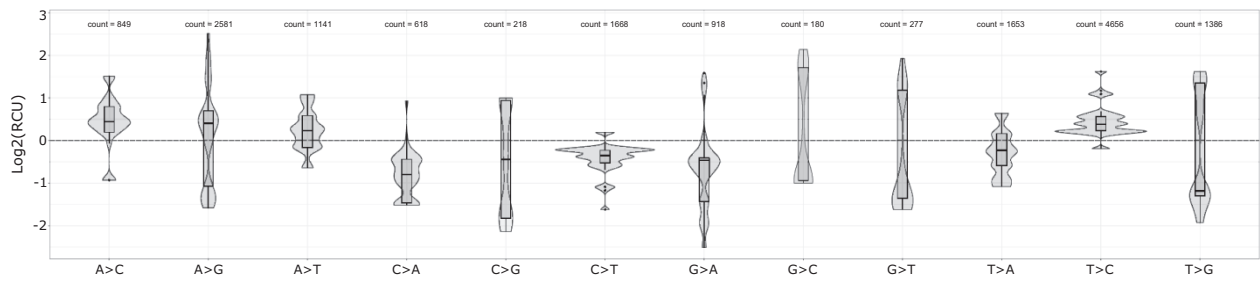


Figure 1. Empirical log₂-transformed RCU distribution of all unique detected substitutions divided by substitution types. Violin and box plots representing the empirical distribution of the log₂-transformed RCU values for each unique variant observed in Dataset 1; mutations observed in multiple samples are considered only once. The mutations were divided into 12 classes related to the possible nucleotide substitutions. For each variant type the overall number of all unique detected substitutions is reported.

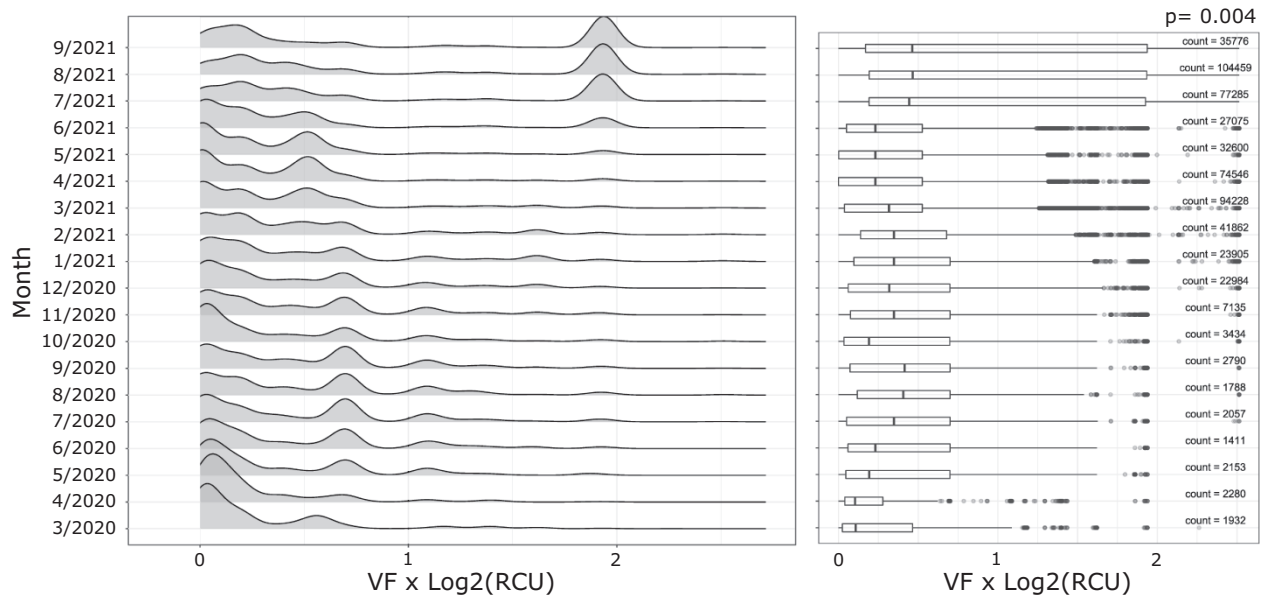


Figure 2. Temporal trend of the log₂-transformed RCU for all the substitutions with positive codon adaptation with respect to the human CU. On the left, we report the distribution of the log₂-transformed RCU multiplied by the VF of every mutation with a positive adaptation to the human CU, binned (per month) with respect to collection date for Dataset 1. On the right, the boxplots of the same distribution are shown and annotated with the number of observations for Dataset 1. The P-value reported on the top right of the Figure is computed performing a Mann-Kendall test for monotonic trend over the mean of the distribution for each month.

we reasoned that the process of synonymous variants selection overtime could be detected in very large time-series analyses. Accordingly, we binned all the silent variants based on their collection date in months intervals and we analyzed the trend of the VF for variants with positive or negative log₂-transformed RCU, where positive values indicate an increase in the similarity to the human CU and negative values the opposite. In particular, we first divided the mutations into two groups, i.e. the ones with a negative RCU (among which we have the majority of the APOBEC-associated mutations) and the ones with a positive RCU. We then grouped variants by collection date (months) and analyzed for each of the two groups the presence of any statistical trend overtime of their RCU multiplied by VF, and finally log₂-transformed. In line with the original hypothesis, silent variants in the positive CU group (Fig. 2 and Supplementary Figs S3–S7) showed an increase of the log₂-transformed VF-weighted RCU over time, significant for all three datasets (one-sided Mann-Kendall test, P-value = 0.004, 0.006, and 0.004 for Dataset 1, Dataset 2, and Dataset 3, respectively), suggesting that CU adaptation might play a significant role in the evolution of SARS-CoV-2 virus. As

expected, no increase could be detected for the variants in the negative CU group. In conclusion, the results provided here point to the evidence that CU adaptation may play a role in the dynamics of minor mutations transitioning to fixed, in addition to functional selection.

Furthermore, to provide additional evidence to confirm that codon adaptation is improving when acquiring mutations over the course of the pandemic, we performed two additional analyses for all the three considered datasets:

- 1) With the aim of evaluating whether, in general, variants leading to better codon adaptation are preferred with respect to the ones not improving CU, we performed the following analysis. We considered the genome position of 100,000 randomly selected mutations (both synonymous and non-synonymous) present in Dataset 1, Dataset 2, and Dataset 3 (e.g. position 150 for variant 150 C>T), tested all the alternative nucleotide substitutions for that position (i.e. 150 C>G and 150 C>A), and checked whether each of such substitutions would lead to the synthesis of the same

protein generated by the original variant (i.e. 150 C>T). This led us to a set of variants where different mutations lead to the synthesis of the same protein. Within this set, we compared the distribution of the CU values obtained with the original variants to the one that would be achieved with the alternative substitutions by performing a standard t-test, which led to a very significant P-value (<0.0001) for all three datasets. When repeating the same analysis with the CoCoPUTs human CU, the results were confirmed.

- 2) We first split the timeline covered by the datasets in two subsequent time frames: (i) first half of the timeline (2020/01/01 to 2020/11/30 for Dataset 1; 2020/01/01 to 2020/10/31 for Dataset 2; 2020/01/01 to 2020/12/31 for Dataset 3), (ii) second half of the timeline (until 2021/09/30 for Dataset 1, 2021/07/31 for Dataset 2 and until 2021/11/30 for Dataset 3); we selected the variants displaying a VF > 0.90 for the first time in either one of the two time frames (i.e. fixed variants); we finally compared the number of such variants belonging to the APOBEC group. Specifically, we reported: for Dataset 1 41.45 per cent of APOBEC variants in time frame (i) and 11.13 in time frame (ii); for Dataset 2 42.34 per cent of APOBEC variants in time frame (i) and 15.17 in time frame (ii); for Dataset 3 38.71 per cent of APOBEC variants in time frame (i) and 16.19 in time frame (ii). We finally performed a standard z-test to compare the proportions for each time frame and obtained a highly significant P-value (<0.0001) for all the three datasets, confirming that APOBEC variants (with negative codon adaptation, on average) show a reduced trend of fixation in the population as the pandemic continues, hinting at the evidence that mutations showing negative CU have lower likelihood of transitioning from minor to fixed.

Taken globally, these results suggest that silent mutations may play a role in viral evolution, by increasing the adaptation of the viral genome to the human CU. To this end, it is tempting to speculate that the observed continuous adaptation of its CU overtime may underlie an increase in the overall efficiency of viral protein production and packaging, with viral genomes with a better adaptation being able to generate more viral particles over time, therefore outperforming other, less adapted, viruses. It is widely accepted that improvements in CU adaptation, by removing translation bottlenecks, may contribute to the overall protein synthesis efficiency (Plotkin and Kudla 2011). Indeed, proteins differing only at silent sites can have several orders of magnitude of difference in their expression level (Kudla et al. 2009). This finding has been widely exploited in both research and industry to adapt coding sequences to specific organisms to optimize protein synthesis. Recently, CU optimization was also proposed as an effective strategy to drive high levels of viral protein expression in human cells to stimulate a very robust immune response (Stachyra et al. 2016). Recent studies pointed also to more complex scenarios, where the use of different codons may allow to fine-tune the translation time of individual gene regions, to allow proper folding of complex domains as well as to maximize the occurrence of post-translational modifications (Fu et al. 2016) in this context a very fast translational process is not always an advantage, as in specific conditions pausing the translation machinery may be beneficial for the production of functional, properly folded proteins. Further analyses and very large datasets will be required to identify these specific events, if present.

As the number of available viral sequences is quickly increasing over time, it will be similarly interesting to isolate individ-

ual silent variants showing relevant changes in their VF over large time spans to study specific CU adaptation patterns in the context of individual viral genes/proteins. Notably, the effect of codon adaptation could be also directly assessed in terms of viral fitness. By using a reverse genetics approach, *in vitro* modified viral RNA genomes could be generated carrying progressively more human-adapted genomes. The functional effect of these modifications could be then assessed in terms of increased viral fitness or competition assays. Similarly, the overall level of protein expression could be tested in target cells by using conventional approaches such as western blot and confocal microscopy.

In conclusion, global as well as time-series analyses of silent mutations indicate that CU adaptation may play a relevant role in the evolution of SARS-CoV-2, suggesting that the evolution of the viral genome has been intense. Further studies will be required to thoroughly dissect the overall impact and clinical relevance of these findings. Soon, it will be important to assess if specific viral genomes with a high CU adaptation also show an increased infectiousness and/or clinical aggressiveness, which, at the time being, is very difficult, owing to the limited availability of clinical data in publicly available SARS-CoV-2 sequencing studies. In this regard and given the profound impact of the COVID-19 pandemic, every effort should be made to share clinical information together with sequencing data, which is unfortunately rarely done.

Data availability

All data are available in the main text or the supplementary materials or can be downloaded from the original SARS-CoV-2 sequence repositories.

Supplementary data

Supplementary data is available at *Virus Evolution* online.

Acknowledgements

We acknowledge that this research was partially supported by the Italian Ministry of University and Research (MIUR)—Department of Excellence project PREMIA (PRECision Medicine Approach: bringing biomarker research to the clinic).

Funding

Italian Ministry of University and Research (MIUR) - Department of Excellence project PREMIA (PRECision Medicine Approach: bringing biomarker research to the clinic). This work was partially supported by a Bicocca 2020 Starting Grant to F.A. and D.R.

Conflict of interest: The founders had no role in the design and conduct of the study; analysis, and interpretation of the data; preparation of the manuscript; and decision to submit the manuscript for publication.

Author contributions

Conceptualization: R.P.; Methodology: R.P., D.R., A.G.; Investigation: D.R., F.A., D.M.; Visualization: F.A., D.M., D.R.; Funding acquisition: R.P., A.G., D.R.; Supervision: R.P., A.G., E.M.E., D.R.; Writing – original draft: D.R., F.A., A.G., R.P. All authors read and approved the final manuscript.

References

- Alexaki, A. et al. (2019) 'Codon and Codon-pair Usage Tables (Cocoputs): Facilitating Genetic Variation Analyses and Recombinant Gene Design', *Journal of Molecular Biology*, 431: 2434–41.
- Alonso, A. M., and Diambra, L. (2020) 'SARS-CoV-2 Codon Usage Bias Downregulates Host Expressed Genes with Similar Codon Usage', *Frontiers in Cell and Developmental Biology*, 8: 831.
- Athey, J. et al. (2017) 'A New and Updated Resource for Codon Usage Tables', *BMC Bioinformatics*, 18: 1–10.
- CDC. (2021), *Centers for Disease Control and Prevention. SARS-CoV-2 Variant Classifications and Definitions* <<https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html>> accessed 21 Mar 2022.
- Chen, F. et al. (2020) 'Dissimilation of Synonymous Codon Usage Bias in Virus–host Coevolution Due to Translational Selection', *Nature Ecology & Evolution*, 4: 589–600.
- Comeron, J. M. (2004) 'Selective and Mutational Patterns Associated with Gene Expression in Humans: Influences on Synonymous Composition and Intron Presence', *Genetics*, 167: 1293–304.
- Dhindsa, R. S. et al. (2020) 'Natural Selection Shapes Codon Usage in the Human Genome', *The American Journal of Human Genetics*, 107: 83–95.
- Duret, L. (2020) 'tRNA Gene Number and Codon Usage in the C. Elegans Genome are Co-adapted for Optimal Translation of Highly Expressed Genes', *Trends in Genetics*, 16: 287–9.
- Forster, P. et al. (2020) 'Phylogenetic Network Analysis of SARS-CoV-2 Genomes', *Proceedings of the National Academy of Sciences*, 117: 9241–3.
- Fu, J. et al. (2016) 'Codon Usage Affects the Structure and Function of the Drosophila Circadian Clock Protein PERIOD', *Genes & Development*, 30: 1761–75.
- Graudenzi, A. et al. (2021) 'Mutational Signatures and Heterogeneous Host Response Revealed via Large-scale Characterization of SARS-CoV-2 Genomic Diversity', *iScience*, 24: 102116.
- Grubaugh, N. D. et al. (2021) 'Public Health Actions to Control New SARS-CoV-2 Variants', *Cell*, 184: 1127–32.
- Ikemura, T. (1981) 'Correlation between the Abundance of Escherichia Coli Transfer RNAs and the Occurrence of the Respective Codons in Its Protein Genes: A Proposal for a Synonymous Codon Choice that Is Optimal for the E. Coli Translational System', *Journal of Molecular Biology*, 151: 389–409.
- (1982) 'Correlation between the Abundance of Yeast Transfer RNAs and the Occurrence of the Respective Codons in Protein Genes: Differences in Synonymous Codon Choice Patterns of Yeast and Escherichia Coli with Reference to the Abundance of Isoaccepting Transfer RNAs', *Journal of Molecular Biology*, 158: 573–97.
- Kanaya, S. et al. (2001) 'Codon Usage and tRNA Genes in Eukaryotes: Correlation of Codon Usage Diversity with Translation Efficiency and with CG-dinucleotide Usage as Assessed by Multivariate Analysis', *Journal of Molecular Evolution*, 53: 290–8.
- Korber, B. et al. (2020) 'Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus', *Cell*, 182: 812–27.
- Kudla, G. et al. (2009) 'Coding-sequence Determinants of Gene Expression in Escherichia Coli', *science*, 324: 255–8.
- Lavner, Y., and Kotlar, D. (2005) 'Codon Bias as a Factor in Regulating Expression via Translation Rate in the Human Genome', *Gene*, 345: 127–38.
- Lemieux, J. E. et al. (2021) 'Phylogenetic Analysis of SARS-CoV-2 in Boston Highlights the Impact of Superspreading Events', *Science*, 371: eabe3261.
- McCarthy, C., Carrea, A., and Diambra, L. (2017) 'Bicodon Bias Can Determine the Role of Synonymous SNPs in Human Diseases', *BMC Genomics*, 18: 1–11.
- Milewska, A. et al. (2018) 'APOBEC3-mediated Restriction of RNA Virus Replication', *Scientific Reports*, 8: 1–12.
- Muik, A. et al. (2021) 'Neutralization of SARS-CoV-2 Lineage B. 1.1. 7 Pseudovirus by BNT162b2 Vaccine-elicited Human Sera', *Science*, 371: 1152–3.
- Nakamura, Y., Gojobori, T., and Ikemura, T. (2000) 'Codon Usage Tabulated from International DNA Sequence Databases: Status for the Year 2000', *Nucleic Acids Research*, 28: 292.
- Plotkin, J. B., and Kudla, G. (2011) 'Synonymous but Not the Same: The Causes and Consequences of Codon Bias', *Nature Reviews Genetics*, 12: 32–42.
- Ramazzotti, D. et al. (2021) 'VERSO: A Comprehensive Framework for the Inference of Robust Phylogenies and the Quantification of Intra-host Genomic Diversity of Viral Samples', *Patterns*, 2: 100212.
- Seemann, T. et al. (2020) 'Tracking the COVID-19 Pandemic in Australia Using Genomics', *Nature Communications*, 11: 1–9.
- Sharp, P. M., Tuohy, T. M. F., and Mosurski, K. R. (1986) 'Codon Usage in Yeast: Cluster Analysis Clearly Differentiates Highly and Lowly Expressed Genes', *Nucleic Acids Research*, 14: 5125–43.
- Stachyra, A. et al. (2016) 'Codon Optimization of Antigen Coding Sequences Improves the Immune Potential of DNA Vaccines against Avian Influenza Virus H5N1 in Mice and Chickens', *Virology Journal*, 13: 1–11.
- Stavrou, S., and Ross, S. R. (2015) 'APOBEC3 Proteins in Viral Immunity', *The Journal of Immunology*, 195: 4565–70.
- Wang, Y. et al. (2021) 'Intra-host Variation and Evolutionary Dynamics of SARS-CoV-2 Populations in COVID-19 Patients', *Genome Medicine*, 13: 1–13.
- WHO. (2022), *WHO Weekly Epidemiological Update* <<https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—8-march-2022>> accessed 21 Mar 2022.
- Williams, T. C., and Burgers, W. A. (2021) 'SARS-CoV-2 Evolution and Vaccines: Cause for Concern?', *The Lancet Respiratory Medicine*, 9: 333–5.
- Woo, P. C. Y. et al. (2007) 'Cytosine Deamination and Selection of CpG Suppressed Clones are the Two Major Independent Biological Forces that Shape Codon Usage Bias in Coronaviruses', *Virology*, 369: 431–42.
- Yang, Z., and Nielsen, R. (1998) 'Synonymous and Nonsynonymous Rate Variation in Nuclear Genes of Mammals', *Journal of Molecular Evolution*, 46: 409–18.
- Zehender, G. et al. (2020) 'Genomic Characterization and Phylogenetic Analysis of SARS-COV-2 in Italy', *Journal of Medical Virology*, 92: 1637–40.