

# Little Evidence the Standard Genetic Code Is Optimized for Resource Conservation

Hana Rozhoňová<sup>1,2</sup> and Joshua L. Payne<sup>\*,1,2</sup>

<sup>1</sup>Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland

<sup>2</sup>Swiss Institute of Bioinformatics, Quartier UNIL-Sorge, Lausanne, Switzerland

\*Corresponding author: E-mail: joshua.payne@env.ethz.ch.

Associate editor: Xuhua Xia

## Abstract

**Selection for resource conservation can shape the coding sequences of organisms living in nutrient-limited environments. Recently, it was proposed that selection for resource conservation, specifically for nitrogen and carbon content, has also shaped the structure of the standard genetic code, such that the missense mutations the code allows tend to cause small increases in the number of nitrogen and carbon atoms in amino acids. Moreover, it was proposed that this optimization is not confounded by known optimizations of the standard genetic code, such as for polar requirement or hydrophathy. We challenge these claims. We show the proposed optimization for nitrogen conservation is highly sensitive to choice of null model and the proposed optimization for carbon conservation is confounded by the known conservative nature of the standard genetic code with respect to the molecular volume of amino acids. There is therefore little evidence the standard genetic code is optimized for resource conservation. We discuss our findings in the context of null models of the standard genetic code.**

**Key words:** translation, standard genetic code, evolutionary design principles.

The standard genetic code (SGC) exhibits numerous optimizations (Freeland et al. 2003). For example, the missense and frameshift mutations allowed by the SGC tend to preserve key physicochemical properties of amino acids, such as polar requirement, hydrophathy, and to a lesser extent, molecular volume (Haig and Hurst 1991, 1999; Geyer and Madany Mamlouk 2018; Wnętrzak et al. 2019; Bartonek et al. 2020; Xu and Zhang 2021a). Recently, an additional optimization was proposed, namely for resource conservation (Shenhav and Zeevi 2020). Motivated by the observation that selection for resource conservation can shape the coding sequences of organisms living in nutrient-limited environments (Mazel and Marlière 1989; Elser et al. 2006; Bragg and Wagner 2007; Lv et al. 2008; Li et al. 2009; Grzymalski and Dussaq 2012; Mende et al. 2017; Hellweger et al. 2018), it was hypothesized that selection for resource conservation has also shaped the structure of the SGC, such that the missense mutations the SGC allows tend to cause small increases in the number of nitrogen and carbon atoms in amino acids. Moreover, it was hypothesized that this optimization is not confounded by known optimizations of the SGC, such as for polar requirement or hydrophathy.

Optimizations in the SGC are typically identified using one of two approaches (Freeland, Knight, and Landweber 2000). In the “engineering approach,” the SGC is compared with codes found by analytical methods or heuristic search algorithms that minimize some objective function, such as the mean absolute change in polar requirement caused by missense mutations (Di Giulio 1989a; Di Giulio et al. 1994; Santos and Monteagudo

2011; Błażej et al. 2018; Wnętrzak et al. 2018). In the “statistical approach,” the SGC is compared with a large number of randomized codes (Alff-Steinberger 1969; Haig and Hurst 1991). The hypothesis that the SGC is optimized for resource conservation was tested using the statistical approach, specifically by quantifying the expected random mutation cost (ERMC) of missense mutations allowed by the SGC, measured in units such as the number of nitrogen or carbon atoms or the absolute change in polar requirement or hydrophathy of amino acids, and comparing this cost to those incurred by 1 million randomized codes (Shenhav and Zeevi 2020).

Because the space of randomized codes is so large ( $\approx 1.5 \times 10^{84}$ ) (Caporaso et al. 2005), it is necessary to draw comparisons with a sample from this space, which can then be used as a null model (Freeland et al. 2003). There are many methods for generating randomized codes, and different methods can generate codes with different properties (Wichmann and Ardern 2019). For example, a method known as quartet shuffling generates randomized codes by shuffling quartet blocks—the blocks of four codons that share the first two bases (e.g., AAA, AAC, AAG, AAU) (Alff-Steinberger 1969; Caporaso et al. 2005). This method, which was used to test the hypothesis that the SGC is optimized for resource conservation (Shenhav and Zeevi 2020), generates randomized codes that preserve two key properties of the SGC, namely the number of codons per amino acid and the degeneracy of the third base (e.g., the three codons for isoleucine always have the same first and second base as the codon for methionine). In contrast, a method known as

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Open Access

amino acid permutation generates randomized codes by permuting the 20 standard amino acids among the synonymous codon blocks (Haig and Hurst 1991). This method, which is most commonly used in the field (Haig and Hurst 1991; Ardell 1998; Freeland and Hurst 1998; Freeland, Knight, Landweber, and Hurst 2000; Gilis et al. 2001; Archetti 2004; Caporaso et al. 2005; Goodarzi, Shateri Najafabadi, Nejad, et al. 2005; Goodarzi, Shateri Najafabadi, and Torabi 2005; Novozhilov et al. 2007; Butler et al. 2009; Tripathi and Deem 2018), generates randomized codes that preserve a different key property of the SGC, namely the structure of the synonymous codons blocks. Importantly, in these randomized codes, the number of codons per amino acid can change drastically relative to the SGC, because the permutation of amino acids among the synonymous blocks is random. These two methods therefore generate randomized codes with substantially different structural properties (supplementary fig. S1, Supplementary Material online).

Here, we test the hypothesis that the SGC is optimized for resource conservation with respect to nitrogen and carbon content by drawing comparisons with randomized codes generated using ten different methods, including quartet shuffling and amino acid permutation (table 1 and supplementary methods and fig. S1, Supplementary Material online). With respect to nitrogen, we find consistent statistical support for resource conservation using only one of the ten methods. With respect to carbon, we find consistent statistical support for resource conservation across the ten methods, but show that this optimization is confounded by the known conservative nature of the SGC with respect to the molecular volume of amino acids (Haig and Hurst 1999).

## Results

### Computing the ERM

We compute the ERM as:

$$\text{ERM} = \sum_{\nu, \nu' \in \mathcal{V}, \nu \neq \nu'} \text{Freq}(\nu) \cdot \text{Prob}(\nu \rightarrow \nu') \cdot \text{Cost}(\nu \rightarrow \nu'), \quad (1)$$

where  $\mathcal{V}$  is the set of all 64 codons,  $\text{Freq}(\nu)$  is the frequency of codon  $\nu$  (supplementary note 1, Supplementary Material

online),  $\text{Prob}(\nu \rightarrow \nu')$  is the probability of mutation from codon  $\nu$  to  $\nu'$  given a genetic code (standard or randomized) and mutation rates (e.g., based on a transition:transversion ratio), and  $\text{Cost}(\nu \rightarrow \nu')$  is the cost of mutating codon  $\nu$  to  $\nu'$  (Shenhav and Zeevi 2020). For resource conservation, the cost is defined as the increase in the number of nitrogen or carbon atoms in the amino acid encoded by codon  $\nu'$  relative to the amino acid encoded by  $\nu$ , whereas for amino acid properties such as polar requirement, hydropathy, and molecular volume, it is defined as the absolute difference in the respective property of the amino acid encoded by codon  $\nu'$  and the amino acid encoded by  $\nu$  (Shenhav and Zeevi 2020).

We use three sets of codon frequencies and mutation rates to compute the ERM of a genetic code (Shenhav and Zeevi 2020) (supplementary methods, Supplementary Material online), namely:

- (1) Baseline parameters: All codon frequencies are equal and mutation rates are based on a transition:transversion ratio of 1:2.
- (2) Ocean parameters: Codon frequencies (supplementary data S1, Supplementary Material online) and mutation rates (supplementary data S2, Supplementary Material online) are derived from marine metagenomics samples (Shenhav and Zeevi 2020).
- (3) Diverse species parameters: Codon frequencies are derived from 39 species (supplementary data S3, Supplementary Material online) (Athey et al. 2017) and mutation rates are based on 11 transition:transversion ratios ranging from 1:5 to 5:1. In total, this set includes 429 combinations of codon frequencies and mutation rates (Shenhav and Zeevi 2020).

For each set, we determine the statistical significance of the ERM of the SGC by computing an empirical  $P$  value, which is the fraction of 1 million randomized genetic codes that have an ERM that is less than or equal to that of the SGC. We compute this empirical  $P$  value separately for each of the ten methods for generating randomized codes. For the “diverse species parameters,” we correct the  $P$  values for testing multiple hypotheses (Benjamini and Hochberg 1995). We report the raw and corrected  $P$  values for all tests in supplementary

**Table 1.** Two Key Properties of the Randomized Genetic Codes Generated with the Ten Different Methods Used in This Study (supplementary methods, Supplementary Material online).

Method	Preserves the Number of Codons per Amino Acid	Preserves the Exact Block Structure of the SGC
Quartet shuffling	Yes	No <sup>a</sup>
Amino acid permutation	No	Yes
Restricted amino acid permutation	No <sup>b</sup>	Yes
N-Block shuffler	Yes	No <sup>a</sup>
Codon shuffler	Yes	No
AAAGALOC shuffler	No	No
Random expansion	No	Yes
Ambiguity reduction 1	No	Yes
Ambiguity reduction 2	No	Yes
2–1–3 model	No	Yes

<sup>a</sup>The randomized codes have a block structure, but it is different from that of the SGC.

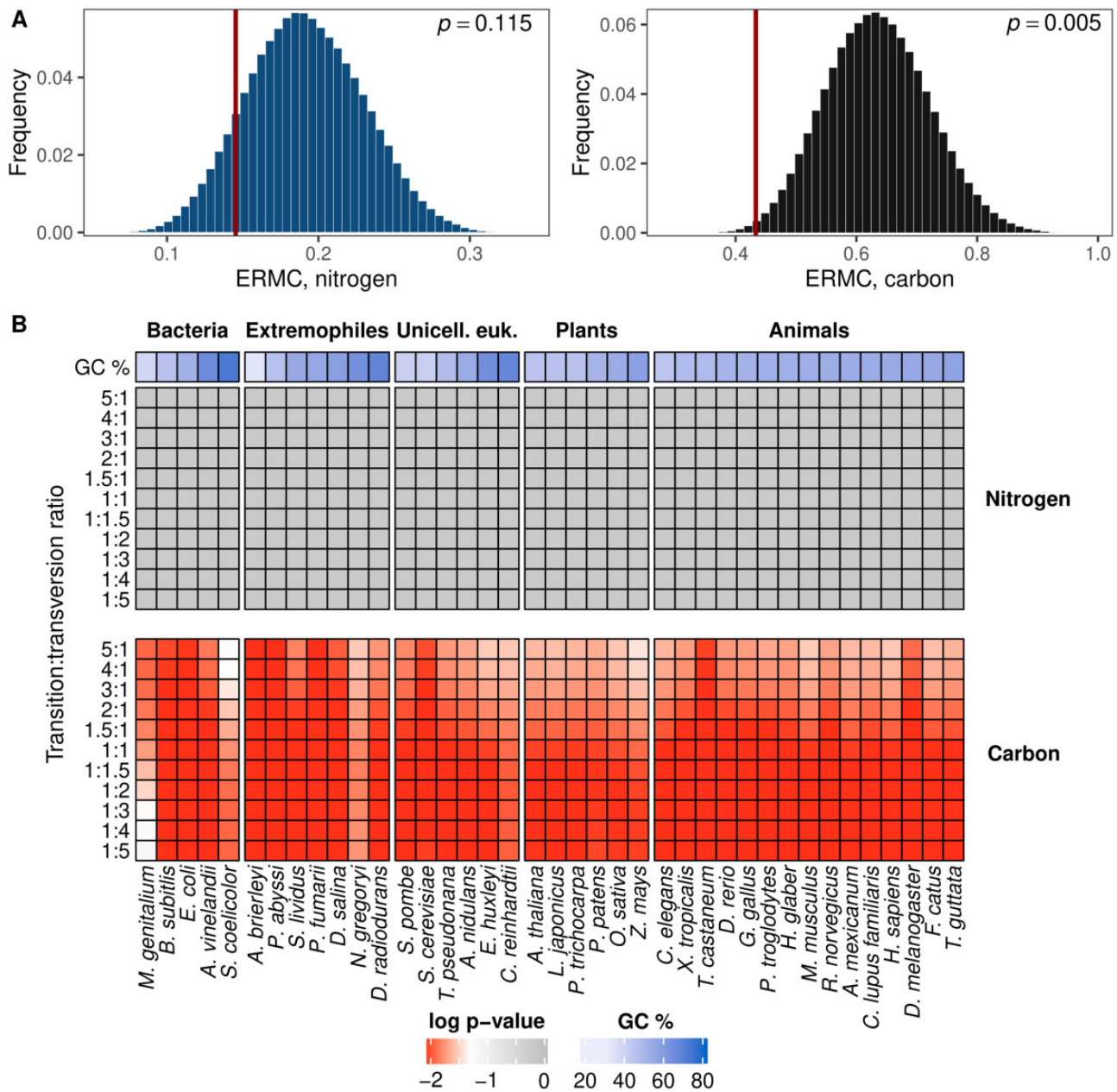
<sup>b</sup>The number of codons per amino acid is allowed to change by at most two, relative to the SGC.

data S4–S12, Supplementary Material online (supplementary table S1, Supplementary Material online).

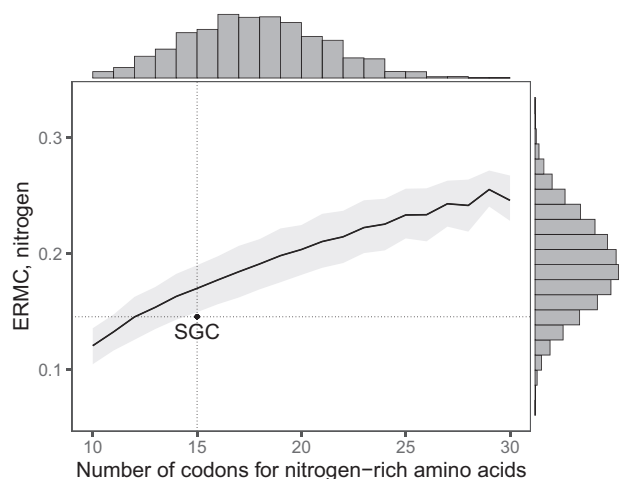
### Nitrogen Conservation Is Highly Sensitive to Choice of Null Model

We find consistent statistical support for nitrogen conservation in the SGC using only one of the ten methods for generating randomized codes (supplementary data S4, Supplementary Material online), namely the codon shuffler (Caporaso et al. 2005) (Baseline parameters:  $P = 1.00 \times 10^{-6}$ ; Ocean

parameters:  $P = 3.00 \times 10^{-6}$ ; Diverse species parameters:  $P \leq 0.016$ ). For the remaining nine methods, nitrogen conservation is never consistently statistically significant across all tested parameters, as illustrated for amino acid permutation in figure 1 (Baseline parameters:  $P = 0.485$ ; Ocean parameters:  $P = 0.115$ ; Diverse species parameters:  $P \geq 0.573$ ). Surprisingly, even for randomized codes generated by quartet shuffling, nitrogen conservation is not statistically significant for the “diverse species parameters” ( $P \geq 0.316$ ), although it is for the “baseline parameters,” ( $P = 0.023$ ) and “ocean parameters” ( $P = 0.034$ ).



**Fig. 1.** Nitrogen conservation is highly sensitive to choice of null model. (A) Histograms of the ERMC for nitrogen (blue) and carbon (black) in 1 million randomized codes generated by amino acid permutation. The vertical red line corresponds to the SGC. Codon frequencies and mutation rates are from the “ocean parameters.” (B)  $P$  values of the ERMC for nitrogen (top) and carbon (bottom) of the SGC, relative to 1 million randomized codes generated by amino acid permutation, using the “diverse species parameters.” Shades of gray correspond to statistically insignificant  $P$  values ( $P > 0.05$ ; darker=less significant) and shades of red to statistically significant  $P$  values ( $P \leq 0.05$ ; darker=more significant). The  $P$  values were adjusted using Benjamini–Hochberg correction for multiple testing. Organisms in each group are ordered based on the GC content of their coding sequences. Unicell. euk., unicellular eukaryotes.



**Fig. 2.** The ERMCM for nitrogen is correlated with the number of codons for nitrogen-rich amino acids. The black line shows the mean, and the shaded area shows the 25th to the 75th quantile, of the ERMCM for nitrogen in relation to the number of codons for nitrogen-rich amino acids in 1 million randomized codes generated by amino acid permutation. The point and dotted lines correspond to the SGC. Histograms of the number of codons for nitrogen-rich amino acids and the ERMCM for nitrogen are shown on the top and on the right of the main panel, respectively. The ERMCM for nitrogen was computed using the “ocean parameters.”

What explains the qualitative difference between the results obtained using these different methods? The key is that the randomized codes generated by both the codon shuffler and quartet shuffling maintain the number of codons per amino acid. In the SGC, the codons of the six nitrogen-rich amino acids (i.e., those with at least one nitrogen atom in their side chain: histidine, lysine, asparagine, glutamine, arginine, and tryptophan) are clustered in the codon table (supplementary fig. S1A and supplementary note 2, Supplementary Material online), such that a point mutation to a codon of a nitrogen-rich amino acid leads with probability 48.9% to a codon of the same or different nitrogen-rich amino acid. Because such clustering is highly unlikely in randomized codes that maintain the number of codons per amino acid, these almost always have a higher ERMCM for nitrogen than the SGC, thus rendering nitrogen conservation statistically significant. In contrast, if the number of codons per amino acid is allowed to change, many randomized codes have a lower ERMCM for nitrogen than the SGC. The reason is that the ERMCM for nitrogen is strongly correlated with the number of codons for nitrogen-rich amino acids in these randomized genetic codes (Pearson’s correlation 0.567,  $P < 2.2 \times 10^{-16}$  for codes generated by amino acid permutation; fig. 2) and this number is often smaller than in the SGC, thus rendering nitrogen conservation statistically insignificant.

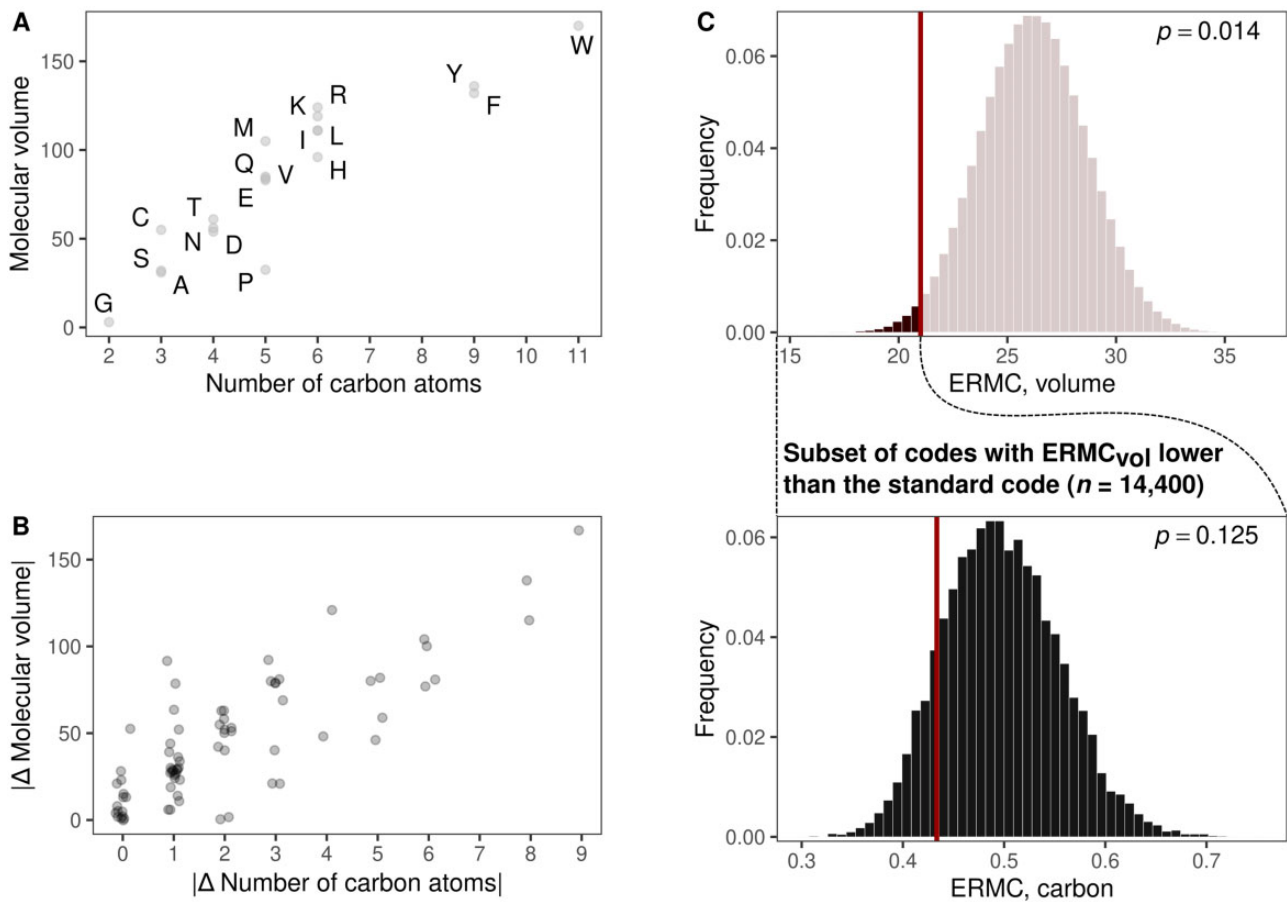
### Carbon Conservation Is Confounded by the Molecular Volume of Amino Acids

We find consistent statistical support for carbon conservation in the SGC across the ten methods for generating randomized codes (Baseline parameters:  $P < 0.05$  for ten of ten methods; Ocean parameters:  $P < 0.05$  for nine of ten methods; Diverse

species parameters: median  $P < 0.05$  for nine of ten methods; fig. 1; supplementary data S5, Supplementary Material online). However, we hypothesize that carbon conservation is confounded by molecular volume (Grantham 1974), because the molecular volume of an amino acid is strongly correlated with its number of carbon atoms (Pearson’s correlation 0.906,  $P = 3.97 \times 10^{-8}$ ; fig. 3A) and the changes caused by missense mutations to amino acids’ molecular volume and number of carbon atoms are therefore strongly correlated (Pearson’s correlation 0.813,  $P < 2.2 \times 10^{-16}$ ; fig. 3B). We test this hypothesis using a hierarchical model (Shenhav and Zeevi 2020). Specifically, for each of the ten methods for generating randomized codes, we examine the subset of randomized codes that have an ERMCM that is less than or equal to that of the SGC for molecular volume and test whether the SGC is also optimized for carbon conservation relative to this subset. It is not (Baseline parameters:  $P > 0.05$  for ten of ten methods; Ocean parameters:  $P > 0.05$  for ten of ten methods; Diverse species parameters: minimum  $P > 0.05$  for nine of ten methods; supplementary data S6, Supplementary Material online), as illustrated in figure 3C for randomized codes generated by amino acid permutation (Baseline parameters:  $P = 0.139$ ; Ocean parameters:  $P = 0.125$ ; Diverse species parameters:  $P > 0.190$ ). Thus, carbon conservation is confounded by the known conservative nature of the SGC with respect to the molecular volume of amino acids (Haig and Hurst 1999).

### Discussion

We found that the proposed optimization of the SGC for nitrogen conservation (Shenhav and Zeevi 2020) is highly sensitive to choice of null model. Specifically, we only found statistical support for nitrogen conservation when using null models that preserve the number of codons per amino acid from the SGC. Choosing an appropriate null model to test for optimizations in the SGC is challenging, because different null models preserve different key properties of the SGC, while perturbing others. Which key properties should be preserved and which should be perturbed? This is a difficult question. On the one hand, null models that preserve the number of codons per amino acid can be justified by correlations between the number of codons per amino acid and the molecular weight of amino acids (Hasegawa and Miyata 1980; Di Giulio 1989b; Dufton 1997) as well as the frequency of amino acids in the proteome (Gilis et al. 2001). However, these correlations are far from perfect (Pearson’s  $R = -0.45$ ,  $P = 0.046$  and  $R = 0.67$ ,  $P = 0.001$ , respectively) and modest changes in the number of codons per amino acid are commonly observed in extant non-standard genetic codes (Knight et al. 2001). On the other hand, null models that preserve the structure of the synonymous codon blocks can be justified by the mode of interaction between mRNA, tRNA, and the ribosome (Ogle et al. 2001, 2003), which results in the third “wobble position” of codons (Crick 1966). However, extant non-standard genetic codes often have synonymous codon blocks that differ from those of the SGC, demonstrating that the exact block structure of



**FIG. 3.** Carbon conservation is confounded by the molecular volume of amino acids. (A) Scatter plot of the number of carbon atoms and the molecular volume of the 20 proteinogenic amino acids. (B) Scatter plot of the absolute change in the number of carbon atoms and the absolute change in molecular volume for the 75 amino acid pairs that are connected by a missense mutation in the SGC. Jitter applied in the x axis for visualization. (C) Histograms of the ERMV for (top) the molecular volume of amino acids in 1 million randomized codes generated by amino acid permutation and (bottom) for carbon in the subset of 14,400 randomized codes that have an ERMV for molecular volume that is less than or equal to that of the SGC. The ERMV was computed using the “ocean parameters.”

the SGC is not the only possible structure (Knight et al. 2001). Given these challenges, a sensible way forward is to use a diversity of null models when testing for optimizations in the SGC (Wichmann and Arderm 2019) and to refrain from reporting optimizations that only find statistical support from a small number of these null models.

Indeed, we found such broad statistical support across a diversity of null models for the proposed optimization for carbon conservation (Shenhav and Zeevi 2020), but we also found that this optimization is confounded by the known conservative nature of the SGC with respect to molecular volume (Haig and Hurst 1999). This highlights another challenge in choosing an appropriate null model to test for optimizations in the SGC: Most null models are agnostic to the evolutionary history of the SGC, which can give the false impression that optimizations are the product of selection rather than a byproduct of the physical processes of gene duplication and mutation (Stoltzfus and Yampolsky 2007; Massey 2008; but, see Di Giulio 2018). Carbon conservation is a case in point. Although there are several nonmutually exclusive hypotheses for how the genetic code evolved (Koonin and Novozhilov 2017), one hypothesis suggests

that in the early stages of code evolution, amino acids were recognized by pockets in the tertiary structure of proto-tRNAs and that the expansion of the code proceeded via duplication and mutation of these proto-tRNAs (Wolf and Koonin 2007). Because a recently duplicated proto-tRNA would likely recognize an amino acid with similar molecular volume to that recognized by its parent proto-tRNA (Massey 2008), gene duplication and mutation would naturally result in a clustering of codons for amino acids with similar molecular volumes in the codon table, as present in the SGC. As carbon is the main building block of all proteinogenic amino acids, the proposed optimization for carbon conservation follows naturally, without needing to evoke selection for resource conservation. More importantly, no matter which model of genetic code evolution one considers, the endpoint is always the SGC, which is conservative with respect to molecular volume (Haig and Hurst 1999). This will always confound carbon conservation.

Finally, we note that if in nutrient-limited environments it is costly for missense mutations to increase the number of nitrogen or carbon atoms in amino acids, then it should be beneficial for missense mutations to decrease the number of

nitrogen or carbon atoms in amino acids. This simple fact makes it difficult to justify the ERM as a measure of the cost of missense mutations, because it only accounts for increases. Indeed, contemporaneous work to ours shows that the SGC is not optimized for resource conservation when the ERM is modified to also account for decreases in the number of nitrogen or carbon atoms (Xu and Zhang 2021b). Taken together, our analyses strongly suggest that the SGC is not optimized for resource conservation.

## Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

## Acknowledgments

We thank Sinisa Bratulic and David M. McCandlish for discussions and feedback on this manuscript. This work was supported by the Swiss National Science Foundation (Grant No. PP00P3\_170604).

## Data Availability

Code used in this study is freely available at <https://github.com/parizkh/resource-conservation-in-genetic-code>.

## References

- Alff-Steinberger C. 1969. The genetic code and error transmission. *Proc Natl Acad Sci U S A*. 64(2):584–591.
- Archetti M. 2004. Codon usage bias and mutation constraints reduce the level of error minimization of the genetic code. *J Mol Evol*. 59(2):258–266.
- Ardell DH. 1998. On error minimization in a sequential origin of the standard genetic code. *J Mol Evol*. 47(1):1–13.
- Athey J, Alexaki A, Osipova E, Rostovtsev A, Santana-Quintero LV, Katneni U, Simonyan V, Kimchi-Sarfaty C. 2017. A new and updated resource for codon usage tables. *BMC Bioinformatics* 18(1):391.
- Bartonek L, Braun D, Zagrovic B. 2020. Frameshifting preserves key physicochemical properties of proteins. *Proc Natl Acad Sci U S A*. 117(11):5907–5912.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Stat Methodol*. 57(1):289–300.
- Błazej P, Wnętrzak M, Mackiewicz D, Mackiewicz P. 2018. Optimization of the standard genetic code according to three codon positions using an evolutionary algorithm. *PLoS One* 13(10):e0205450.
- Bragg JG, Wagner A. 2007. Protein carbon content evolves in response to carbon availability and may influence the fate of duplicated genes. *Proc Biol Sci*. 274(1613):1063–1070.
- Butler T, Goldenfeld N, Mathew D, Luthey-Schulten Z. 2009. Extreme genetic code optimality from a molecular dynamics calculation of amino acid polar requirement. *Phys Rev E*. 79:060901.
- Caporaso JG, Yarus M, Knight RD. 2005. Error minimization and coding triplet/binding site associations are independent features of the canonical genetic code. *J Mol Evol*. 61(5):597–607.
- Crick F. 1966. Codon–anticodon pairing: the wobble hypothesis. *J Mol Biol*. 19(2):548–555.
- Di Giulio M. 1989a. The extension reached by the minimization of the polarity distances during the evolution of the genetic code. *J Mol Evol*. 29(4):288–293.
- Di Giulio M. 1989b. Some aspects of the organization and evolution of the genetic code. *J Mol Evol*. 29(3):191–201.
- Di Giulio M. 2018. A non-neutral origin for error minimization in the origin of the genetic code. *J Mol Evol*. 86(9):593–597.
- Di Giulio M, Capobianco M, Medugno M. 1994. On the optimization of the physicochemical distances between amino acids in the evolution of the genetic code. *J Theor Biol*. 168(1):43–51.
- Dufton MJ. 1997. Genetic code synonym quotas and amino acid complexity: cutting the cost of proteins? *J Theor Biol*. 187(2):165–173.
- Elser JJ, Fagan WF, Subramanian S, Kumar S. 2006. Signatures of ecological resource availability in the animal and plant proteomes. *Mol Biol Evol*. 23(10):1946–1951.
- Freeland SJ, Hurst LD. 1998. The genetic code is one in a million. *J Mol Evol*. 47(3):238–248.
- Freeland SJ, Knight RD, Landweber LF. 2000. Measuring adaptation within the genetic code. *Trends Biochem Sci*. 25(2):44–45.
- Freeland SJ, Knight RD, Landweber LF, Hurst LD. 2000. Early fixation of an optimal genetic code. *Mol Biol Evol*. 17(4):511–518.
- Freeland SJ, Wu T, Keulmann N. 2003. The case for an error minimizing standard genetic code. *Orig Life Evol Biosph*. 33(4–5):457–477.
- Geyer R, Madany Mamlouk A. 2018. On the efficiency of the genetic code after frameshift mutations. *PeerJ* 6:e4825.
- Gilis D, Massar S, Cerf NJ, Rooman M. 2001. Optimality of the genetic code with respect to protein stability and amino-acid frequencies. *Genome Biol*. 2(11):research0049.1.
- Goodarzi H, Shateri Najafabadi H, Nejad HA, Torabi N. 2005. The impact of including tRNA content on the optimality of the genetic code. *Bull Math Biol*. 67(6):1355–1368.
- Goodarzi H, Shateri Najafabadi H, Torabi N. 2005. On the coevolution of genes and genetic code. *Gene* 362:133–140.
- Grantham R. 1974. Amino acid difference formula to help explain protein evolution. *Science* 185(4154):862–864.
- Grzymalski JJ, Dussaq AM. 2012. The significance of nitrogen cost minimization in proteomes of marine microorganisms. *ISME J*. 6(1):71–80.
- Haig D, Hurst LD. 1991. A quantitative measure of error minimization in the genetic code. *J Mol Evol*. 33(5):412–417.
- Haig D, Hurst LD. 1999. A quantitative measure of error minimization in the genetic code. *J Mol Evol*. 49(5):708.
- Hasegawa M, Miyata T. 1980. On the antisymmetry of the amino acid code table. *Orig Life*. 10(3):265–270.
- Hellweger FL, Huang Y, Luo H. 2018. Carbon limitation drives GC content evolution of a marine bacterium in an individual-based genome-scale model. *ISME J*. 12(5):1180–1187.
- Knight RD, Freeland SJ, Landweber LF. 2001. Rewiring the keyboard: evolvability of the genetic code. *Nat Rev Genet*. 2(1):49–58.
- Koonin EV, Novozhilov AS. 2017. Origin and evolution of the universal genetic code. *Annu Rev Genet*. 51:45–62.
- Li N, Lv J, Niu DK. 2009. Low contents of carbon and nitrogen in highly abundant proteins: evidence of selection for the economy of atomic composition. *J Mol Evol*. 68(3):248–255.
- Lv J, Li N, Niu DK. 2008. Association between the availability of environmental resources and the atomic composition of organismal proteomes: evidence from *Prochlorococcus* strains living at different depths. *Biochem Biophys Res Commun*. 375(2):241–246.
- Massey SE. 2008. A neutral origin for error minimization in the genetic code. *J Mol Evol*. 67(5):510–516.
- Mazel D, Marlière P. 1989. Adaptive eradication of methionine and cysteine from cyanobacterial light-harvesting proteins. *Nature* 341(6239):245–248.
- Mende DR, Bryant JA, Aylward FO, Eppley JM, Nielsen T, Karl DM, DeLong EF. 2017. Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat Microbiol*. 2(10):1367–1373.
- Novozhilov AS, Wolf YI, Koonin EV. 2007. Evolution of the genetic code: partial optimization of a random code for robustness to translation error in a rugged fitness landscape. *Biol Direct*. 2:24.
- Ogle JM, Brodersen DE, Clemons WM, Tarry MJ, Carter AP, Ramakrishnan V. 2001. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 292(5518):897–902.
- Ogle JM, Carter AP, Ramakrishnan V. 2003. Insights into the decoding mechanism from recent ribosome structures. *Trends Biochem Sci*. 28(5):259–266.

- Santos J, Monteagudo A. 2011. Simulated evolution applied to study the genetic code optimality using a model of codon reassignments. *BMC Bioinformatics* 12:56.
- Shenhav L, Zeevi D. 2020. Resource conservation manifests in the genetic code. *Science* 370(6517):683–687.
- Stoltzfus A, Yampolsky LY. 2007. Amino acid exchangeability and the adaptive code hypothesis. *J Mol Evol.* 65(4):456–462.
- Tripathi S, Deem MW. 2018. The standard genetic code facilitates exploration of the space of functional nucleotide sequences. *J Mol Evol.* 86(6):325–339.
- Wichmann S, Ardern Z. 2019. Optimality in the standard genetic code is robust with respect to comparison code sets. *Biosystems* 185:104023.
- Wnętrzak M, Błażej P, Mackiewicz D, Mackiewicz P. 2018. The optimality of the standard genetic code assessed by an eight-objective evolutionary algorithm. *BMC Evol Biol.* 18(1):192.
- Wnętrzak M, Błażej P, Mackiewicz P. 2019. Optimization of the standard genetic code in terms of two mutation types: point mutations and frameshifts. *Biosystems* 181:44–50.
- Wolf YI, Koonin EV. 2007. On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biol Direct.* 2:14.
- Xu H, Zhang J. 2021a. On the origin of frameshift-robustness of the standard genetic code. *Mol Biol Evol.* doi: 10.1093/molbev/msab164.
- Xu H, Zhang J. 2021b. Is the genetic code optimized for resource conservation? *Mol Biol Evol.* doi: 10.1093/molbev/msab239.