



Cite this article: Chiyomaru K, Takemoto K. 2020 Revisiting the hypothesis of an energetic barrier to genome complexity between eukaryotes and prokaryotes. *R. Soc. open sci.* **7**: 191859.
<http://dx.doi.org/10.1098/rsos.191859>

Received: 23 October 2019
Accepted: 21 January 2020

Subject Category:
Genetics and genomics

Subject Areas:
bioinformatics/evolution/microbiology

Keywords:
metabolic rate, genome size,
gene number, cell size, evolution

Author for correspondence:
Kazuhiro Takemoto
e-mail: takemoto@bio.kyutech.ac.jp

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.4840296>.

Revisiting the hypothesis of an energetic barrier to genome complexity between eukaryotes and prokaryotes

Katsumi Chiyomaru and Kazuhiro Takemoto

Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Iizuka, Fukuoka 820-8502, Japan

 KT, 0000-0002-6355-1366

The absence of genome complexity in prokaryotes, being the evolutionary precursors to eukaryotic cells comprising all complex life (the prokaryote–eukaryote divide), is a long-standing question in evolutionary biology. A previous study hypothesized that the divide exists because prokaryotic genome size is constrained by bioenergetics (prokaryotic power per gene or genome being significantly lower than eukaryotic ones). However, this hypothesis was evaluated using a relatively small dataset due to lack of data availability at the time, and is therefore controversial. Accordingly, we constructed a larger dataset of genomes, metabolic rates, cell sizes and ploidy levels to investigate whether an energetic barrier to genome complexity exists between eukaryotes and prokaryotes while statistically controlling for the confounding effects of cell size and phylogenetic signals. Notably, we showed that the differences in bioenergetics between prokaryotes and eukaryotes were less significant than those previously reported. More importantly, we found a limited contribution of power per genome and power per gene to the prokaryote–eukaryote dichotomy. Our findings indicate that the prokaryote–eukaryote divide is hard to explain from the energetic perspective. However, our findings may not entirely discount the traditional hypothesis; in contrast, they indicate the need for more careful examination.

1. Introduction

Eukaryotic cells arose from prokaryotes and are comprised of all complex life. Biological complexity (e.g. genome complexity, cellular complexity and multicellularity) is believed to harbour several advantages. Genome size and the number of genes (i.e. genome complexity) increase with environmental variability because organisms need more functional (e.g. metabolic) genes to

adapt to changing environments (e.g. nutrient variability) [1,2]. Cellular complexity may enhance biological robustness [3], and multicellular organisms have evolved sophisticated, higher-level functionality via cooperation among component cells with complementary behaviours [4,5]. However, only some prokaryotes have evolved biological complexity. The large gap between prokaryotes and eukaryotes (the prokaryote–eukaryote divide) is a long-standing mystery in evolutionary biology [6–8].

Lane & Martin focused on genome complexity and hypothesized that the prokaryote–eukaryote divide is due to the prokaryotic genome size being constrained by bioenergetics [6] (Lane–Martin hypothesis). Their report is positioned as a proposal of a hypothesis rather than a data analysis; however, they used the data on genome size, metabolic rate (oxygen consumption rate) and ploidy level of 12 (cellular) eukaryotes and 55 prokaryotes, to demonstrate the validity of their hypothesis. In particular, Lane & Martin showed that the power (i.e. the oxygen available) per gene and power per genome of eukaryotes was significantly larger (approximately 2000-fold) than those of prokaryotes. This indicates the presence of an energetic barrier against genome complexity between prokaryotes and eukaryotes. They concluded that eukaryotes have allowed the expansion of their genome sizes via endosymbiosis, giving rise to mitochondria, which have provided an energetic boost.

However, the Lane–Martin hypothesis is controversial. For example, Lynch [9] pointed out that the increase in genome complexity can be explained through non-adaptive evolutionary processes. Booth and Doolittle argued that eukaryogenesis—the crossing of the deep gulf between prokaryotes and eukaryotes—lacks rigorous evidential and statistical support [10]. The Lane–Martin hypothesis has several limitations. Primarily, the hypothesis was based on a biased evaluation in a limited number of species due to lack of data availability at the time. In addition, the hypothesis was based on the metabolic rate of prokaryotes grown in the presence of various substrates (i.e. under nutrient-rich conditions). The use of such metabolic rates as a measure of power production may not be informative from an evolutionary perspective [11,12].

Moreover, the effects of cell mass were not statistically controlled. Metabolic rate has a strong positive correlation with body mass (cell mass in the case of cellular organisms); in particular, the relationship between metabolic rate and body mass approximately obeys a power law [13–16]. Lynch & Marinov [11,12] investigated a common currency of energy per unit of cell volume and found no energetic difference between eukaryotes and prokaryotes. This finding eliminates the need to invoke an energetics barrier hypothesis to genome complexity; however, it was based on a biased evaluation in a limited number of species. More importantly, the effects of phylogenetic signals were not considered, although the importance of phylogeny in evaluating associations between biological features has been well-established through comparative phylogenetic analyses [17,18]. An opposite conclusion may be derived when considering comparative phylogenetic analysis [19,20].

Therefore, in this study, we revisited the Lane–Martin hypothesis. In particular, a larger dataset of genomes, metabolic rates, ploidy level and cell sizes (masses) was constructed, and the contribution of energetic parameters to prokaryote/eukaryote classification (prokaryote–eukaryote divide) was investigated, while statistically controlling for the potentially confounding effect of cell sizes. Comparative phylogenetic analyses were also performed to evaluate the effects of phylogenetic signals on the contribution of energetic parameters to the prokaryote–eukaryote dichotomy.

2. Material and methods

2.1. Metabolic rate and cell mass

Based on a previous study [6], we collected data on mass-specific metabolic rates and cell mass of prokaryotes and eukaryotes (protozoa) from the literature [21,22]. Additionally, we used Supporting Datasets S1 (heterotrophic prokaryotes), S2 (heterotrophic protozoa), S7 (cyanobacteria) and S8 (eukaryotic microalgae) from a previous study [16] to obtain additional data on mass-specific metabolic rates (oxygen consumption rate) and cell masses (electronic supplementary material, dataset S1). The units of mass-specific metabolic rates and cell masses were converted to watt per kilogram ($W\text{ kg}^{-1}$) and picograms (pg), respectively. For a species, multiple values of mass-specific metabolic rates may be available in the dataset. For a comparison with the previous study [6], despite criticism by Lynch & Marinov [11,12], the maximum mass-specific metabolic rates were used to estimate energy supply (electronic supplementary material, dataset S1); specifically, they mainly correspond to mass-specific metabolic rates measured at the exponential or logarithmic growth phase and summit metabolic rates. Moreover, cell mass associated with the metabolic rate was used.

2.2. Genome size, gene number and ploidy level

We selected prokaryotic and eukaryotic species whose complete genomes were available in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [23] and/or National Center for Biotechnology Information (NCBI) database. Haploid genome sizes (bp) and the number of total protein-coding genes (haploid gene number) of these species were obtained from the databases according to species names (electronic supplementary material, dataset S1). When multiple strains for a species were available in the databases, we selected one strain as a representative of the species according to the year in which its genome was first completely determined. Data were not available for some species, and in these instances, the genome size and gene number of a different species within the same genus whose genome was available in the database were used. Specifically, the substitute genome was selected from a different species in the same genus according to the year in which its genome was first completely determined. The data on genome size and gene number were downloaded from the KEGG database on 2 April 2018 and the NCBI database on 12 December 2018, respectively.

Following a previous study [6], we also collected data on the ploidy level (electronic supplementary material, dataset S1). For eukaryotes, ploidy levels were retrieved from the literature according to species names. For prokaryotes, ploidy levels were retrieved from the literature according to species names because some bacteria may be oligoploid and polyploid [24,25]; however, it was assumed that prokaryotes whose ploidy level had not been reported in any previous studies were monoploid. Prokaryotes are generally assumed to be monoploid during slow growth [24]; moreover, the increase in ploidy levels, observed when bacteria grow fast, are transient. For some species, ploidy levels of a different species within the same genus were used because species-specific ploidy levels were unavailable in the dataset. For a given species, multiple values of ploidy levels may be available. In the analyses, the maximum ploidy level was used, which mainly corresponds to the ploidy level observed at the exponential or logarithmic growth phase for each species.

Finally, we obtained a larger dataset of genome, power and cell sizes for 36 eukaryotes and 156 prokaryotes.

2.3. Energetic parameters

Following a previous study [6], we used the data on mass-specific metabolic rate (B_c), cell mass (M), haploid genome size (G), haploid gene number (N_g) and ploidy level (P) to calculate the following energetic parameters: power per cell ($B_c \times M$; fW), power per haploid genome ($1000 \times \text{power per cell}/G/P$; pW), power per gene (power per cell/ N_g/P ; fW) and power per genome (power per gene $\times N_g = \text{power per cell}/P$; fW). The primary focus was on power per genome and power per gene for comparison with the previous study.

2.4. Data analyses

All statistical tests were performed using R software (version 3.6.1; www.R-project.org).

To evaluate the contribution of the energetic parameters and cell size (mass) to the prokaryote–eukaryote dichotomy (or divide), logistic regression analyses were conducted using R software. No biological replicates in the dataset were used in the analyses. The energetic parameters and cell masses were log-transformed. The quantitative variables were normalized to the same scale, with a mean of 0 and a standard deviation of 1, using the scale function in R before the analysis.

To remove the effects of phylogenetic signals from the regression analyses, phylogenetic logistic regression analyses [26,27] were performed using the function *binaryPGLMM* in the R-package *ape* (version 5.3). In this function, s_2 is the scaling component of the variance in the model, where $s_2 = 0$ suggests no phylogenetic signal, and a high s_2 value implies a strong phylogenetic signal [28]. The phylogenetic tree, required for phylogenetic regression, was constructed using conserved protein-coding genes, following a previous study [29]. The conserved genes were determined based on the KEGG Orthology (KO) database. We selected 17 eukaryotes and 122 prokaryotes available in the KO database and used 12 KO groups conserved in these organisms for phylogenetic tree construction (electronic supplementary material, table S1). The sequences of genes in these groups were downloaded from the KEGG database on January 17, 2019 and were aligned using MULTiple Sequence Comparison by Log-Expectation (MUSCLE; version 3.8.31) [30] with the parameter ‘*-maxiterate 1000*’ and the resulting alignments were processed using the Gblocks program (version 0.91b) [31] with the default settings to eliminate poorly aligned positions. The processed alignments were concatenated and subjected to phylogenetic analysis. The phylogenetic tree was constructed using Molecular Evolutionary Genetics Analysis (MEGA; version 7) software [32]. We performed model selection based on the Akaike

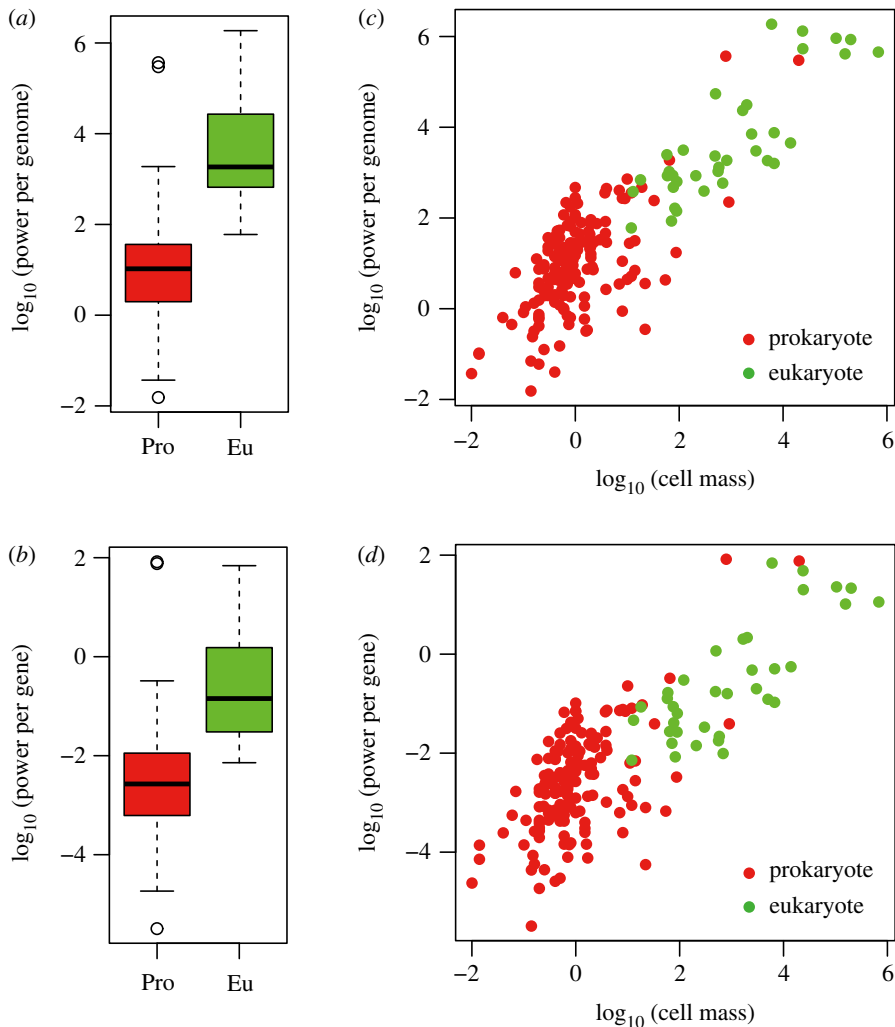


Figure 1. Comparison of energetic measures between prokaryotes and eukaryotes. (a) The difference in power per genome between prokaryotes (Pro) and eukaryotes (Eu). (b) The difference in power per gene between prokaryotes (Pro) and eukaryotes (Eu). (c) Scatter plot of power per genome versus cell mass. (d) Scatter plot of power per gene and cell mass. Power per genome, power per gene and cell mass were base 10 log-transformed.

information criterion (AIC) value. The substitution LG+G model, the best-fit model, was employed to produce the maximum-likelihood tree (electronic supplementary material, dataset S1 and figure S1).

The contribution (i.e. non-zero estimate) of each explanatory variable to the prokaryote–eukaryote dichotomy was considered significant when the associated p -value was less than 0.05. We used the function *R2.pred* in R package *rr2* (version 1.0.2) [28] to calculate the coefficients of determination of the standard and phylogenetic logistic regression models.

3. Results

The increased genome complexity in eukaryotes was re-confirmed. The haploid genome sizes (44 Mbp in median) of eukaryotes were larger than that (4 Mbp in median) of prokaryotes ($p < 2.2 \times 10^{-16}$ using the Wilcoxon test). The haploid gene number (14973 in median) of eukaryotes was greater than that (3935 in median) of prokaryotes ($p < 2.2 \times 10^{-16}$ using the Wilcoxon test).

Inspired by the previous study of Lane & Martin [6], we investigated the differences in metabolic power between prokaryotes and eukaryotes and re-confirmed that the metabolic power of eukaryotes was greater than that of prokaryotes. Specifically, the power per cell (4819 fW in median) of eukaryotes was greater than that (1.7 fW in median) of prokaryotes ($p < 2.2 \times 10^{-16}$ using the Wilcoxon test). More importantly, the power per genome (1850 fW in median) of eukaryotes was larger than that (1 fW in median) of prokaryotes (figure 1a; $p < 2.2 \times 10^{-16}$ using the Wilcoxon test); in addition,

Table 1. Contributions of explanatory variables to the prokaryote–eukaryote dichotomy. The variable ‘dichotomy’ indicates whether a species is a eukaryote (1) or not (0). s.e. and AIC correspond to the standard error and Akaike information criterion value of the model, respectively. R^2 is the coefficient of determination.

model	variable	estimate	s.e.	p -value	AIC	R^2
(a) dichotomy \sim power per genome + cell mass	power per genome	0.62	0.65	0.34	65.3	0.73
	cell mass	2.88	0.71	5.1×10^{-5}		
(b) dichotomy \sim power per gene + cell mass	power per gene	−0.18	0.49	0.71	67.2	0.70
	cell mass	3.46	0.67	2.6×10^{-7}		
(c) dichotomy \sim power per genome	power per genome	3.21	0.56	1.3×10^{-8}	87.3	0.62
(d) dichotomy \sim power per gene	power per gene	2.14	0.38	1.9×10^{-8}	118.1	0.42
(e) dichotomy \sim cell mass	cell mass	3.33	0.57	3.9×10^{-9}	65.3	0.71

the power per haploid genome size (3.7×10^{-2} pW in median) of eukaryotes was larger than that (2.3×10^{-3} pW in median) of prokaryotes ($p = 1.5 \times 10^{-10}$ using the Wilcoxon test). Moreover, the power per gene (0.14 fW in median) of eukaryotes was larger than that (0.0027 fW in median) of prokaryotes (figure 1b; $p = 3.1 \times 10^{-15}$ using the Wilcoxon test).

However, differences in cell mass between prokaryotes and eukaryotes were observed; in particular, the cell mass (570 pg median) of eukaryotes was larger than that (0.7 pg in median) of prokaryotes ($p < 2.2 \times 10^{-16}$ using the Wilcoxon test). The distribution of the cell mass of prokaryotes was partially overlapped with that of eukaryotes. For example, the cell mass, power per genome, and power per gene of several prokaryotes (e.g. *Thioploca* and *Trichodesmium* species) almost equalled those of eukaryotes. Moreover, the power per genome and power per gene for prokaryotes appeared to be similar to eukaryotes for a similar cell mass. This indicates that the contributions of power per genome and power per gene in the prokaryote–eukaryote dichotomy depend on cell mass. In particular, the power per genome (figure 1c) and power per gene (figure 1d) were co-related to cell mass in a linear fashion within and between prokaryotic and eukaryotic groups.

Therefore, a standard logistic regression analysis was performed to statistically control for the effects of cell mass; specifically, multivariate regression models were constructed encompassing an energetic parameter and cell mass. The regression analyses showed that the contributions of power per genome (table 1a) and power per gene (table 1b) to the prokaryote–eukaryote dichotomy were not statistically significant. Instead, the cell mass was a dominant indicator for the dichotomy (eukaryotic cells are larger than prokaryotic cells). For comparison, single regression analyses (table 1c–e) were also performed. For power per genome, the AIC values for the single regression model (table 1c) were higher than that of the multivariate regression model (table 1a). For power per gene, the AIC value of the single model (table 1d) was also higher than that of the multivariate model (table 1b). These results indicate that cell mass, not power per genome and power per gene, account for the prokaryote–eukaryote dichotomy. In the single regression models, the AIC value of the model for cell mass is the smallest (tie). The result also indicates that cell mass predominately contributes to the prokaryote–eukaryotes dichotomy (table 1e).

However, there is a possibility of phylogenetic signals affecting conclusions obtained from the standard regression analyses. Therefore, phylogenetic logistic regression analyses were performed to remove the phylogenetic effects. High s^2 values indicated the importance of phylogenetic signals (table 2). The single phylogenetic regression models indicated that the contribution of power per genome (table 2c) to the prokaryote–eukaryote dichotomy was less significant, contrary to the results of the standard logistic regression analyses (table 1). The difference in power per gene (table 2d) between prokaryotes and eukaryotes was observed; however, this difference was not exceedingly statistically significant. The cell mass of eukaryotes was larger than that of prokaryotes (table 2e); however, the contribution of cell mass was not exceedingly significant. Multivariate regression models were also constructed to statistically control for the effect of cell mass. The regression models also showed that the contributions of power per genome (table 2a) and power per gene (table 2b) to the prokaryote–eukaryote dichotomy were not statistically significant; moreover, cell mass also contributed slightly to the dichotomy.

Table 2. Contributions of explanatory variables to the prokaryote–eukaryote dichotomy when removing the effects of phylogenetic signals. The variable ‘dichotomy’ indicates whether a species is a eukaryote (1) or not (0). s.e. corresponds to the standard error. s2 indicates a phylogenetic signal (see §2.4). Values in brackets are the associated p -value. R^2 is the coefficient of determination.

model	variable	estimate	s.e.	p -value	s2	R^2
(a) dichotomy \sim power per genome + cell mass	power per genome	0.06	1.46	0.96	1.38 (1.1×10^{-4})	0.98
	cell mass	1.56	1.37	0.29		
(b) dichotomy \sim power per gene + cell mass	power per gene	−0.48	1.13	0.67	1.40 (1.2×10^{-4})	0.99
	cell mass	2.09	1.40	0.14		
(c) dichotomy \sim power per genome	power per genome	1.43	0.77	0.065	1.37 (1.8×10^{-5})	0.99
(d) dichotomy \sim power per gene	power per gene	1.05	0.67	0.117	1.35 (2.9×10^{-6})	0.99
(e) dichotomy \sim cell mass	cell mass	1.61	0.76	0.035	1.35 (1.1×10^{-4})	0.98

4. Discussion

These results indicate no difference in power per genome and power per gene between prokaryotes and eukaryotes, which is not consistent with Lane & Martin’s conclusion that the prokaryotic genome size is constrained by bioenergetics. The simple comparison tests (figure 1*a,b*) indicated that the power per genome and median power per gene of eukaryotes were greater than those of prokaryotes; however, the observed differences were artefacts due to no consideration of the effects of cell mass and phylogeny. The result that cell size (mass) showed a linear relationship with power per genome (figure 1*c*) and power per gene (figure 1*d*) indicates a lack of difference in power per genome and power per gene between prokaryotes and eukaryotes for similar cell mass. Standard and phylogenetic logistic regression analyses (tables 1*a,b* and 2*a,b*) showed no contribution of power per genome and power per gene to the prokaryote–eukaryote dichotomy when statistically controlling for the effect of cell size. Moreover, no difference in power per genome (table 2*c*) and power per gene between (table 2*d*) prokaryotes and eukaryotes was observed even if the effect of cell mass was not statistically controlled and disregarding the effects of phylogenetic signals. The results indicate that there is slight difference in power per genome and power per gene at the root of the phylogenetic tree and that a Brownian motion-like evolution could explain the differences in power per genome (figure 1*c*) and power per gene (figure 1*d*) observed (i.e. at the leaf level of the tree).

The observed differences in power per genome and power per gene between eukaryotes and prokaryotes were less significant than those previously reported. Specifically, the power per genome of eukaryotes was approximately 1850-fold greater ($=1850/1$) than that of prokaryotes, although Lane and Martin reported that the power per genome of eukaryotes was approximately 10 000-fold greater ($=1143/0.12$; see Table 1 in [6]) than that of prokaryotes. Moreover, the power per gene of eukaryotes was approximately 52-fold greater ($=0.14/0.0027$) than that of prokaryotes, although the previous study reported that the power per gene of eukaryotes was roughly 200-fold greater ($=57.15/0.03$; see Table 1 in [6]). This discrepancy might be due to differences in the datasets and data analyses between this study and the previous study. In this study, the data on metabolic rate, cell mass and genome were collected from more 36 eukaryotes and 156 prokaryotes, whereas the previous study only considered 12 eukaryotes and 55 prokaryotes. The dataset in this study partly overlaps with the dataset used in the previous study of Lane & Martin [6] because the same literature [21,22] was used. However, we were not able to substantially evaluate how our dataset was different from the dataset used in the previous study by Lane & Martin [6]. We requested the original dataset from one of the authors. However, we were informed that the dataset was currently unavailable; in particular, the author needed to relocate literature sources for genome size, ploidy, metabolic rates, etc., as the original study was much older; as of October 2018. A comparison of datasets may be evaluated in the near future.

The findings of this study are inconsistent with the idea that cells with greater internal complexity impose greater energy supply (i.e. Lane–Martin hypothesis). The findings indicate the prokaryote–eukaryote divide is harder to explain than previously thought; rather, they support the hypothesis of the passive emergence of genome complexity by non-adaptive processes [9,11,12]. As Lynch & Marinov [12] mentioned, the origin of the mitochondrion was not a prerequisite for genome-size expansion, although the origin was a key event in evolutionary history (e.g. the acquisition of eukaryote-specific traits such as the cell cycle, sex, phagocytosis, endomembrane trafficking, the nucleus and multicellularity [6,33]); rather, genome-size expansion passively occurred in species experiencing relatively low efficiency of selection due to small effective population sizes. Koonin [34] also stated that eukaryotic cells emerged at least in part by initial non-adaptive processes made possible due to a strong and prolonged population bottleneck.

The definition of power per genome and power per gene is still a matter of controversy. The conclusion in this study is limited to power per genome and power per gene, as defined in Lane & Martin’s original study [6]. As Lynch & Marinov [11,12] pointed out, the use of metabolic rate may not be helpful as a measure of power production, as it may fail to distinguish between the investment in cellular reproduction and that associated with non-growth-related processes (e.g. diversity of cellular functions, ranging from turnover of biomolecules, intracellular transport, control of osmotic balance and membrane potential, nutrient uptake, information processing, and motility). Therefore, more suitable measures are needed for more careful examination. For example, Lynch & Marinov used the number of ATP → ADP turnovers as a common currency of energy and found that the costs of a gene at the DNA, RNA and protein levels declines with cell volume in both bacteria and eukaryotes, relative to the lifetime ATP requirements of a cell. However, Lane & Martin [35] stated that the number of ATP → ADP turnovers is not an alternative measure of the power per gene [6] because it corresponds to energy demand whereas power per gene [6] is considered as energy availability per gene, i.e. supply, not demand. This discrepancy is due to Lane & Martin stating in their original study that power per gene represents the cost of expressing the gene [6]. Lynch & Marinov [36] also pointed out this misleading expression. To avoid this dissonance, a more explicit and easily assessed definition of power per gene may be needed. For example, it may be useful to consider genes in a specific functional category [37]. In this study, the power per gene was used based on the metabolic rate because we aimed to revisit the Lane–Martin hypothesis [6], and the amount of available data on the number of ATP → ADP turnovers was limited. However, as with Lynch & Marinov, our study supports the conclusion that power per gene hardly contributes to the prokaryote–eukaryote divide. In addition, the study of Lynch & Marinov [11] was criticized in terms of elusive data and reproducibility [38]; however, this was caused by the authors’ failure to note the citations for these data [39]. This indicates that access to open data is also important for debate in the prokaryote–eukaryote divide.

The current study has several limitations. Only organisms for which complete genome sequences were available were considered in order to accurately estimate power per genome and power per gene and in order to perform the phylogenetic comparative analysis. The findings of this study depend significantly on the quality of genome annotation. Moreover, as previously mentioned [40], there are limitations to the phylogenetic comparative analysis. This type of analysis assumes a Brownian motion-like evolution of biological traits on a phylogenetic tree with accurate branch lengths, which may lead to a misleading conclusion. For example, statistical power decreases when a dataset is reduced in size following phylogenetic corrections [41]. In particular, the dataset used in this study contained only a few samples for eukaryotes. Therefore, the continued sequencing of genomes from a wide range of organisms is important.

The ploidy level is still controversial, although data were collected on ploidy in organisms as much as possible; however, we assumed that prokaryotes whose ploidy level had not been reported in any previous studies were monoploid. This limitation may not pose a problem because bacteria are generally assumed to be monoploid [24] and the increase in ploidy levels, observed when bacteria grow fast, are transient. Moreover, similar tendencies (i.e. limited contributions of power per genome and power per gene to the prokaryote–eukaryote dichotomy) were observed in standard regression analyses even when prokaryotes whose ploidy level had not been reported were removed (electronic supplementary material, table S2). However, the cost of polyploidy should be considered because prokaryotes are believed to require extreme polyploidy to scale up to the eukaryotic size [6,8,35,42]. The prokaryotic ploidy level was correlated with cell mass (Spearman’s rank correlation coefficient $r_s = 0.43$ and the associated p -value $p = 0.0092$) in the dataset, where the prokaryotes whose ploidy level had not been reported were removed. It may be necessary to consider extreme polyploid bacteria such as *Thiomargarita* and *Epulopiscium*, which have multiple copies of their genome [6,8,35,42], although the dataset in this study included prokaryotes with relatively high ploidy levels (highest value = 218). However, it was

necessary to exclude these species in the data analyses because the parameters required were unavailable and/or ambiguous. For example, accurately annotated genomes are required for calculating power per genome and power per gene and for performing phylogenetic analyses. However, the *Thiomargarita* genome was not complete. The data on the *Epulopiscium* metabolic rate was unavailable. The *Thiomargarita* ploidy level was ambiguous; in particular, it was only retrieved from personal communication [6]. To avoid this limitation, the genomes of more extreme polyploid prokaryotes need to be completed. Moreover, ploidy levels in more organisms need to be identified using real-time polymerase chain reaction (PCR) methods [24,43] because the ploidy level may not be conserved within the same phylogenetic groups, and there may be no obvious correlations between the ploidy levels with primary parameters (e.g. haploid genome size and mode of life) [24].

In conclusion, the findings of this study indicate no energetic barrier to genome complexity between prokaryotes and eukaryotes, contrary to the Lane–Martin hypothesis [6]. Despite the limitations in our data analyses, our findings advance our understanding of the energetics of genome complexity and the prokaryote–eukaryote divide. However, these findings may not entirely discount the traditional hypothesis; instead, they indicate the requirement for a more careful examination using more comprehensive analyses. In particular, this study emphasizes the importance of rigorous evidential and statistical support for debate in the prokaryote–eukaryote divide.

Data accessibility. The datasets supporting this article have been uploaded as electronic supplementary material.

Authors' contributions. K.T. conceived and designed the study. K.C. and K.T. prepared the data. K.C. and K.T. performed data analysis and interpreted the results. K.C. and K.T. drafted the manuscript. All authors have read and approved the final manuscript.

Competing interests. The authors declare no competing interests.

Funding. This study was supported by a Grant-in-Aid for Young Scientists (A) from the Japan Society for the Promotion of Science (grant no. 17H04703).

Acknowledgements. The authors are much obliged to Prof. William F. Martin and Dr Hideto Takami for providing their useful comments on bacterial ploidy. The authors would like to thank Editage (www.editage.com) for English language editing.

References

- Sabath N, Ferrada E, Barve A, Wagner A. 2013 Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biol. Evol.* **5**, 966–977. (doi:10.1093/gbe/evt050)
- Bentkowski P, Van Oosterhout C, Mock T. 2015 A model of genome size evolution for prokaryotes in stable and fluctuating environments. *Genome Biol. Evol.* **7**, 2344–2351. (doi:10.1093/gbe/evv148)
- Stelling J, Sauer U, Szallasi Z, Doyle FJ, Doyle J. 2004 Robustness of cellular functions. *Cell* **118**, 675–685. (doi:10.1016/j.cell.2004.09.008)
- Ratcliff WC, Denison RF, Borrello M, Travisano M. 2012 Experimental evolution of multicellularity. *Proc. Natl Acad. Sci. USA* **109**, 1595–1600. (doi:10.1073/pnas.1115323109)
- Kirk DL. 2005 A twelve-step program for evolving multicellularity and a division of labor. *Bioessays* **27**, 299–310. (doi:10.1002/bies.20197)
- Lane N, Martin W. 2010 The energetics of genome complexity. *Nature* **467**, 929–934. (doi:10.1038/nature09486)
- Lane N. 2011 Energetics and genetics across the prokaryote–eukaryote divide. *Biol. Direct* **6**, 35. (doi:10.1186/1745-6150-6-35)
- Lane N. 2014 Bioenergetic constraints on the evolution of complex life. *Cold Spring Harb. Perspect. Biol.* **6**, a015982. (doi:10.1101/cshperspect.a015982)
- Lynch M. 2007 The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl Acad. Sci. USA* **104**, 8597–8604. (doi:10.1073/pnas.0702207104)
- Booth A, Doolittle WF. 2015 Eukaryogenesis, how special really? *Proc. Natl Acad. Sci. USA* **112**, 10 278–10 285. (doi:10.1073/pnas.1421376112)
- Lynch M, Marinov GK. 2017 Membranes, energetics, and evolution across the prokaryote–eukaryote divide. *Elife* **6**, 1–30. (doi:10.7554/eLife.20437)
- Lynch M, Marinov GK. 2015 The bioenergetic costs of a gene. *Proc. Natl Acad. Sci. USA* **112**, 201514974. (doi:10.1073/pnas.1514974112)
- Brown JH, Gillooly JF, Allen AP, Savage VM, West GB. 2004 Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789. (doi:10.1890/03-9000)
- Speakman JR. 2005 Body size, energy metabolism and lifespan. *J. Exp. Biol.* **208**, 1717–1730. (doi:10.1242/jeb.01556)
- West GB, Woodruff WH, Brown JH. 2002 Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. *Proc. Natl Acad. Sci. USA* **99**, 2473–2478. (doi:10.1073/pnas.012579799)
- Makarieva AM, Gorshkov VG, Li B-L, Chown SL, Reich PB, Gavrillo VM. 2008 Mean mass-specific metabolic rates are strikingly similar across life's major domains: evidence for life's metabolic optimum. *Proc. Natl Acad. Sci. USA* **105**, 16 994–16 999. (doi:10.1073/pnas.0802148105)
- Garland T, Bennett AF, Rezende EL. 2005 Phylogenetic approaches in comparative physiology. *J. Exp. Biol.* **208**, 3015–3035. (doi:10.1242/jeb.01745)
- Garland T, Harvey PH, Ives AR. 1992 Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst. Biol.* **41**, 18–32. (doi:10.1093/sysbio/41.1.18)
- Takemoto K, Yoshitake I. 2013 Limited influence of oxygen on the evolution of chemical diversity in metabolic networks. *Metabolites* **3**, 979–992. (doi:10.3390/metabo3040979)
- Naisbit RE, Kehrl P, Rohr RP, Bersier L-F. 2011 Phylogenetic signal in predator–prey body-size relationships. *Ecology* **92**, 2183–2189. (doi:10.1890/10-2234.1)
- Makarieva AM, Gorshkov VG, Li B-L. 2005 Energetics of the smallest: do bacteria breathe at the same rate as whales? *Proc. R. Soc. B* **272**, 2219–2224. (doi:10.1098/rspb.2005.3225)
- Fenchel T, Finlay BJ. 1983 Respiration rates in heterotrophic, free-living protozoa. *Microb. Ecol.* **9**, 99–122. (doi:10.1007/BF02015125)
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016 KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462. (doi:10.1093/nar/gkv1070)
- Pecoraro V, Zerulla K, Lange C, Soppa J. 2011 Quantification of ploidy in proteobacteria

- revealed the existence of monoploid, (mero-) oligoploid and polyploid species. *PLoS ONE* **6**, e16392. (doi:10.1371/journal.pone.0016392)
25. Mendell JE, Clements KD, Choat JH, Angert ER. 2008 Extreme polyploidy in a large bacterium. *Proc. Natl Acad. Sci. USA* **105**, 6730–6734. (doi:10.1073/pnas.0707522105)
 26. Ives AR, Garland T. 2010 Phylogenetic logistic regression for binary dependent variables. *Syst. Biol.* **59**, 9–26. (doi:10.1093/sysbio/syp074)
 27. Ives AR, Garland T. 2014 Phylogenetic regression for binary dependent variables. In *Modern phylogenetic comparative methods and their application in evolutionary biology* (ed. LZ Garamszegi), pp. 231–261. Berlin, Germany: Springer.
 28. Nakagawa S, Schielzeth H. 2013 A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* **4**, 133–142. (doi:10.1111/j.2041-210x.2012.00261.x)
 29. Takami H, Arai W, Takemoto K, Uchiyama I, Taniguchi T. 2015 Functional classification of uncultured ‘*Candidatus* Caldiarchaeum subterraneum’ using the Maple system. *PLoS ONE* **10**, e0132994. (doi:10.1371/journal.pone.0132994)
 30. Edgar RC. 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797. (doi:10.1093/nar/gkh340)
 31. Talavera G, Castresana J. 2007 Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577. (doi:10.1080/10635150701472164)
 32. Kumar S, Stecher G, Tamura K. 2016 MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874. (doi:10.1093/molbev/msw054)
 33. Martin WF. 2017 Symbiogenesis, gradualism, and mitochondrial energy in eukaryote evolution. *Period. Biol.* **119**, 141–158. (doi:10.18054/pb.v119i3.5694)
 34. Koonin EV. 2015 Energetics and population genetics at the root of eukaryotic cellular and genomic complexity. *Proc. Natl Acad. Sci. USA* **112**, 15 777–15 778. (doi:10.1073/pnas.1520869112)
 35. Lane N, Martin WF. 2016 Mitochondria, complexity, and evolutionary deficit spending. *Proc. Natl Acad. Sci. USA* **113**, E666. (doi:10.1073/pnas.1522213113)
 36. Lynch M, Marinov GK. 2016 Reply to Lane and Martin: Mitochondria do not boost the bioenergetic capacity of eukaryotic cells. *Proc. Natl Acad. Sci. USA* **113**, E667–E668. (doi:10.1073/pnas.1523394113)
 37. Takemoto K, Kawakami Y. 2015 The proportion of genes in a functional category is linked to mass-specific metabolic rate and lifespan. *Sci. Rep.* **5**, 10008. (doi:10.1038/srep10008)
 38. Gerlitz M, Knopp M, Kapust N, Xavier JC, Martin WF. 2018 Elusive data underlying debate at the prokaryote-eukaryote divide. *Biol. Direct* **13**, 21. (doi:10.1186/s13062-018-0221-x)
 39. Lynch M, Marinov GK. 2018 Response to Martin and colleagues: Mitochondria do not boost the bioenergetic capacity of eukaryotic cells. *Biol. Direct* **13**, 9–10. (doi:10.1186/s13062-018-0228-3)
 40. Takemoto K, Imoto M. 2017 Exosomes in mammals with greater habitat variability contain more proteins and RNAs. *R. Soc. open sci.* **4**, 170162. (doi:10.1098/rsos.170162)
 41. Griffith OL, Moodie GEE, Civetta A. 2003 Genome size and longevity in fish. *Exp. Gerontol.* **38**, 333–337.
 42. Lane N. 2017 Serial endosymbiosis or singular event at the origin of eukaryotes? *J. Theor. Biol.* **434**, 58–67. (doi:10.1016/j.jtbi.2017.04.031)
 43. Griese M, Lange C, Soppa J. 2011 Ploidy in cyanobacteria. *FEMS Microbiol. Lett.* **323**, 124–131. (doi:10.1111/j.1574-6968.2011.02368.x)