ARTICLE

Check for updates

Obligate endosymbiosis enables genome expansion during eukaryogenesis

Samuel H. A. von der Dunk [1✉], Paulien Hogeweg[1] & Berend Snel [1]

The endosymbiosis of an alpha-proteobacterium that gave rise to mitochondria was one of the key events in eukaryogenesis. One striking outcome of eukaryogenesis was a much more complex cell with a large genome. Despite the existence of many alternative hypotheses for this and other patterns potentially related to endosymbiosis, a constructive evolutionary model in which these hypotheses can be studied is still lacking. Here, we present a theoretical approach in which we focus on the consequences rather than the causes of mitochondrial endosymbiosis. Using a constructive evolutionary model of cell-cycle regulation, we find that genome expansion and genome size asymmetry arise from emergent host–symbiont cell-cycle coordination. We also find that holobionts with large host and small symbiont genomes perform best on long timescales and mimic the outcome of eukaryogenesis. By designing and studying a constructive evolutionary model of obligate endosymbiosis, we uncovered some of the forces that may drive the patterns observed in nature. Our results provide a theoretical foundation for patterns related to mitochondrial endosymbiosis, such as genome size asymmetry, and reveal evolutionary outcomes that have not been considered so far, such as cell-cycle coordination without direct communication.

---

[1] Theoretical Biology and Bioinformatics, Department of Biology, Science Faculty, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.
✉email: s.h.a.vonderdunk@uu.nl

Mitochondrial endosymbiosis was one of the pivotal steps in the evolution of eukaryotes from prokaryotes[1]. Yet, the sequence of events during eukaryogenesis, let alone their driving forces, remains largely unknown due to the lack of intermediate species between prokaryotes and eukaryotes and the lack of evolutionary trajectories analogous to eukaryogenesis. Currently, conflicting theories have been proposed about the role of mitochondrial endosymbiosis in the complexification of the host genome and cell. The mito-late hypothesis posits that complexity was required for endosymbiosis, such as phagocytotic machinery to engulf the endosymbiont[2], a nucleus to protect the host genome against foreign molecules[3], and complex regulation and intracellular trafficking systems to control and communicate with endosymbionts[4]. Conversely, the mito-early hypothesis states that mitochondria were required for complexity to provide the energy for a large genome and cell[5–7].

The mitochondrial endosymbiosis step in eukaryogenesis is a pre-eminent example of a transition in individuality[8], where two (or more, e.g., López-García and Moreira[9]) prokaryotes lost their autonomy, and the proto-eukaryotic holobiont became the main unit of evolution[10,11]. Models are especially powerful for probing big questions where data are limited. In particular, constructive modeling, in which multiple levels and multiple degrees of freedom are incorporated, has been successful in providing insights into the emergence of complexity[12], e.g., in the RNA world[13] and at the origin of multicellularity[14].

Most hypotheses regarding the role of mitochondrial endosymbiosis in eukaryogenesis are focused on the causes of endosymbiosis, in particular the ecological and metabolic context, but many controversies remain (e.g., Zachar and Szathmáry[15]; Box 1 in López-García and Moreira[9]). However, the metabolic repertoire contributed relatively little to eukaryotic complexity[16]—as opposed to the regulatory repertoire, for instance—so a pure metabolic viewpoint is insufficient to understand the relation between endosymbiosis and complexity. Here we want to explore a complementary viewpoint, looking at the consequences rather than the causes of endosymbiosis by studying how obligate endosymbiosis between two prokaryote-like entities shapes the evolution of their genomes and gene regulatory networks. By designing and studying a constructive evolutionary model of cell-cycle regulation, we address the question of how endosymbiosis and complexity are related. Apart from this general question, we study with our model more specific hypotheses about the evolutionary mechanisms underlying eukaryotic complexity.

One hallmark of biological complexity that theories on eukaryogenesis seek to explain is the large eukaryotic genome. At least part of this eukaryotic genome originates from endosymbiotic gene transfer and is therefore a consequence of endosymbiosis. Several driving forces have been proposed to explain rampant endosymbiotic gene transfer from the mitochondrion to the nucleus, but the validity and quantitative contribution of each of these forces remain unresolved. For instance, endosymbiotic gene transfer could grant the host tighter regulatory control over the symbiont[17] or rescue symbiont genes from Muller's ratchet caused by high mutation rate, population bottlenecks and lack of recombination[18,19]. Alternatively, numeric dominance of the endosymbiotic genome might be sufficient to explain how most genes end up in the nucleus through a non-adaptive ratchet-like process, except for a few special cases[20].

In addition to gene transfer from the mitochondrion, gene duplication was an important source of eukaryotic complexity[16]. What drives such host complexification is unclear, in particular, because no intuition can be derived from present-day endosymbiotic relations which involve a host that is already a complex eukaryote. Proto-eukaryotes have been hypothesized to live in smaller populations than prokaryotes which would have weakened selection for fast growth and thereby allowed genome expansion[21]. In line with the theory of Constructive Neutral Evolution[22], the complexity of some large molecular machines appears to be redundant[23,24]. Generally, however, the complex structures and behaviors that arise in eukaryotes are considered to be beneficial.

For the symbiont, gene loss through endosymbiotic gene transfer was supplemented by gene deletion. Once an organism is evolutionarily confined inside a host, Muller's ratchet can explain genome shrinkage[25]. Additionally, adaptive forces may drive genome shrinkage by reducing the cost of genome maintenance and replication. The photosynthetic endosymbiont of *Paulinella chromatophora*, which was acquired through a relatively recent primary endosymbiosis event, has lost numerous genes that are essential for autonomous growth[26]. Similarly, intracellular parasites are hypothesized to have small genomes because the relatively constant host environment does not demand behavioral complexity[27]. Thus, depending on the details of the endosymbiotic relation, both adaptive and non-adaptive forces are likely to play a role in the genomic streamlining of endosymbionts.

Results

To investigate how endosymbiosis impacts the evolution of complexity, we designed a multilevel model which is simple yet contains various degrees of freedom to allow self-organization through evolution. Autonomous cells, which carry their own genome and execute their own cell cycle, are forced into an obligate endosymbiosis with one host and one or more symbionts. Hosts and symbionts implement a non-linear genotype-phenotype map where mutations occur at the genome level. Cell behavior, i.e., replication and division, is directly linked to the expression of cell-cycle genes, viz. *Caulobacter crescentus*[28]. Thus, the model does not specify an explicit fitness criterion, which facilitates the emergence of unforeseen evolutionary patterns. Below, the model is explained in more detail.

Cell-cycle regulation in hosts and symbionts. To study endosymbiosis separate from the metabolic context, we model the autonomous regulation of host and symbiont cell cycles using our recently published model (Fig. 1a, top; Von der Dunk et al.[29]). In this model, a gene regulatory network arising from stochastic interactions between genes and binding sites on a genome generates cyclic expression dynamics that represent a cell cycle. Rather than being evaluated with an explicit fitness criterion, cells have to execute the correct cell-cycle stages after which they divide. We model the replication of the genome explicitly, and replication speed—the stretch of genome replicated in one time step—is given by the nutrient abundance, which in turn depends on preset influx rate into the environment and local population density. In particular, replication occurs in the S-stage, which should therefore be visited sufficiently many times to allow the entire genome to be replicated. Thus replication speed is directly linked to genome size, and small genomes are favored because they allow cells to execute faster cell cycles. Through genome replication, the genome organization also feeds back to the regulatory level as genes are replicated at different times during the cell cycle, and gene copy numbers influence binding probabilities.

Evolution of complex cell-cycle regulation in prokaryotes. Before describing the extended endosymbiosis model, we now briefly summarize the modeling results from the evolution of cell-cycle regulation in free-living cells representing prokaryotes[29]. In general, the tight link between genome size and growth rate constrains the evolution of complexity in prokaryotes. Yet despite
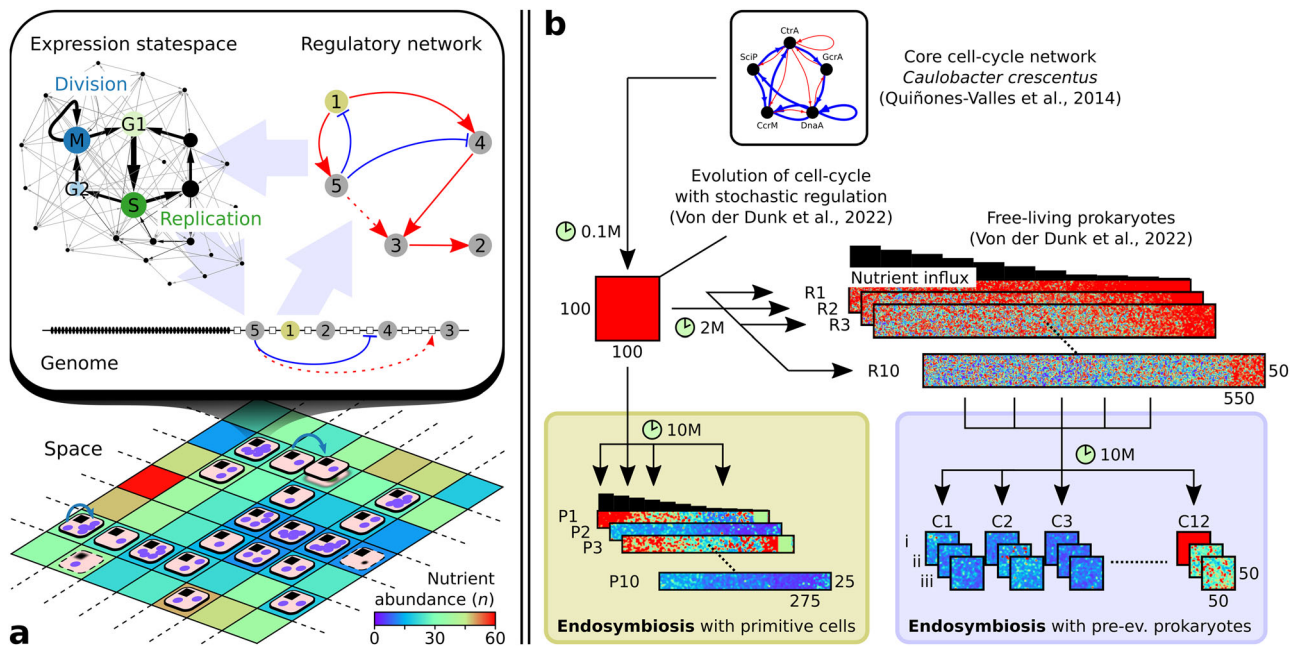
Fig. 1 Overview of the model and experimental setup. a Host and symbiont each consist of a genome (with discrete beads representing genes and binding sites), a gene regulatory network and a cell cycle. Holobionts (disks) live in space and comprise one host (black square) and one or more symbionts (purple circles) that compete for nutrients. For both the host and symbiont, a division event (blue arrow) and a death event (blurred shape) are shown. Lattice site colors depict nutrient abundance for an influx of $n_{influx} = 30$. **b** Two in silico evolution experiments with endosymbiosis were performed (green and purple boxes): one starting with primitive cells on a nutrient gradient and one starting with pre-evolved prokaryotes from our previous work[29]. Simulation time and grid sizes are shown for each experiment; colors depict nutrient abundance.

strong selection for small genomes, some populations achieve adaptation to harsh conditions through genome expansion which allows for complex cell-cycle behavior. Furthermore, cells exploited and augmented the effect of genome replication on gene regulation in order to adapt to fluctuations in nutrient conditions without having to expand their regulatory repertoire. In four of ten evolutionary replicates, cells evolved a de novo cell-cycle checkpoint allowing them to time their cell cycle with nutrient abundance, thus behaving as generalists. In the other replicates, cellular adaptation to different nutrient conditions was to varying extent achieved by speciation, resulting in multiple specialist strains along the nutrient gradient.

Constructive model of obligate endosymbiosis. We model obligate endosymbiosis by forcing two cells, as described above, into one holobiont (Fig. 1a, bottom). Herein the holobiont is defined as a higher-level entity that comprises a single host and one or more symbionts cf. refs. [10,30]. The holobiont dies when the host dies or when the last symbiont dies, which can be due to random death ($d = 0.001$) or due to reaching the M-stage prematurely, i.e., without having finished genome replication or without having expressed all preceding cell-cycle stages. The holobiont divides when the host divides by reaching M-stage after a complete cell cycle. The new holobiont then overgrows one of the neighboring sites, killing the previous occupant, if any, and inherits each symbiont of the parental holobiont with a probability of 0.5. Due to the stochastic inheritance of symbionts, obligate endosymbiosis forces holobionts to maintain at least two and likely more symbionts at division time to produce viable offspring.

Besides their obligate relationship, hosts and symbionts compete for nutrients in the environment. At each site, nutrients that flux in are equally divided over all hosts and symbionts in the 3-by-3 neighborhood. Through nutrient depletion, high symbiont numbers lead to very slow replication rates, which can cause

stagnation of the host cell cycle, eventually killing the holobiont. In sum, holobionts require neither too few nor too many symbionts to be both stable (ensuring that offspring receives at least one symbiont) and competitive (ensuring fast growth and preventing host starvation).

Genomes of real organisms encode many genes that are not related to cell-cycle regulation, such as metabolic genes. In our model, each genome carries 50 household genes that are required to perform other tasks for the host or symbiont. To study hypotheses on genome size evolution, we allow these household functions to be encoded by either the host or the symbionts or by both, requiring only that they together encode at least 100 household genes (wherein those of the symbionts are averaged). Through duplication and deletion, household genes can be transferred indirectly between host and symbiont, which, along with changes in regulatory repertoire sizes, allows for the evolution of a large host and small symbiont genome or vice versa.

Eukaryogenesis from primitive or complex FECA. To study the cellular and genomic impact of obligate endosymbiosis, we performed two in silico evolution experiments initialized with holobionts of different complexity, which we term primitive and complex FECA, inspired by the First Eukaryotic Common Ancestor (Fig. 1b). For the first experiment, we selected 12 different host–symbiont pairs (C1–12; complex FECA) from the final successful free-living prokaryotes that evolved in Von der Dunk et al.[29] (see Supplementary Tables S1 and S2). For each pair, we ran three technical replicates (i–iii), allowing us to determine how much these initial hosts and symbionts influenced the outcome of evolution. Holobionts adapted quickly and often successfully to the endosymbiotic condition, as hosts and symbionts were already capable of executing complex cell-cycle behavior (Supplementary Figs. S4 and S5). Yet the pre-evolved regulatory topologies were not very amenable to changes, so cells

turned out to be restricted in their ways of adapting to endo-symbiosis, as shown below.

For the second experiment, we evolved cell-cycle behavior from scratch (P1–10; primitive FECA), i.e., starting with the most primitive cell-cycle network that was also used to initiate evolution in our previous study (Von der Dunk et al.[29], derived from *Caulobacter crescentus* in Quiñones-Valles et al.[28]). As in those previous experiments, evolution was carried out on a nutrient gradient where cells could adapt to different nutrient conditions. Since initial regulation was very primitive, cell-cycle adaptation could be accomplished in multiple ways yielding substantial variation in evolutionary trajectories between replicate experiments. Yet adaptation was slow, and most populations adapted poorly compared to the evolution experiments with pre-evolved cell-cycle behavior. Nevertheless, we will focus on the replicates starting with a primitive FECA (P1–10) to disentangle how cell behavior has adapted to endosymbiosis and then turn our focus to the replicates starting with a complex FECA (C1–12) to understand how genome size evolution interacts with endosymbiosis.

Holobionts adapt to nutrient gradient. Fig. 2a shows the adaptive process of the 10 in silico evolution replicates starting with primitive hosts and symbionts (P1–10; see "Methods"). In three replicates, the small initial population was unable to improve cell-cycle behavior and stabilize the holobiont, resulting in early extinction before $t = 10^6$. Yet in the seven remaining replicates, holobionts managed

to adapt to limited nutrient conditions and the endosymbiotic life-style, reaching final population sizes ranging from about 28% (P6) to 63% (P9) of the grid. Thus, the experiment has successfully generated diverse evolutionary trajectories emanating from the same origin and experiencing the same environmental conditions.

Populations grow by expanding their range on the nutrient gradient and by increasing their density. In the most successful replicate (P9), we can distinguish two adaptive phases ($t \approx 10^6$ and $5 \cdot 10^6 < t < 8 \cdot 10^6$) when the population expands its range to poorer conditions and simultaneously increases its density in rich conditions (Fig. 3). In general, we find that adaptation occurs in relatively short phases interspersing long periods of stasis (Fig. 2a, viz. Gould and Eldredge[31]), which is similar to what we found in our previous model that represents free-living prokaryotes[29].

Genome expansion and adaptation. In all replicates, host and symbiont genomes expand rapidly early in evolution (Fig. 2b, c). Large genomes are *prima facie* unfavorable because they take long to replicate, forcing cells to execute slow cell cycles. Interestingly, the rapid expansion occurs when holobionts are still primitive and unfit, at population sizes below 12% of the grid ($N < 825$). The three replicates that went extinct experienced genome expansion during their entire existence. Early genome expansion thus apparently results from a lack of selection. At the same time, expanding the regulatory repertoire allows hosts and symbionts to invent more complex cell-cycle behavior and thereby underlies
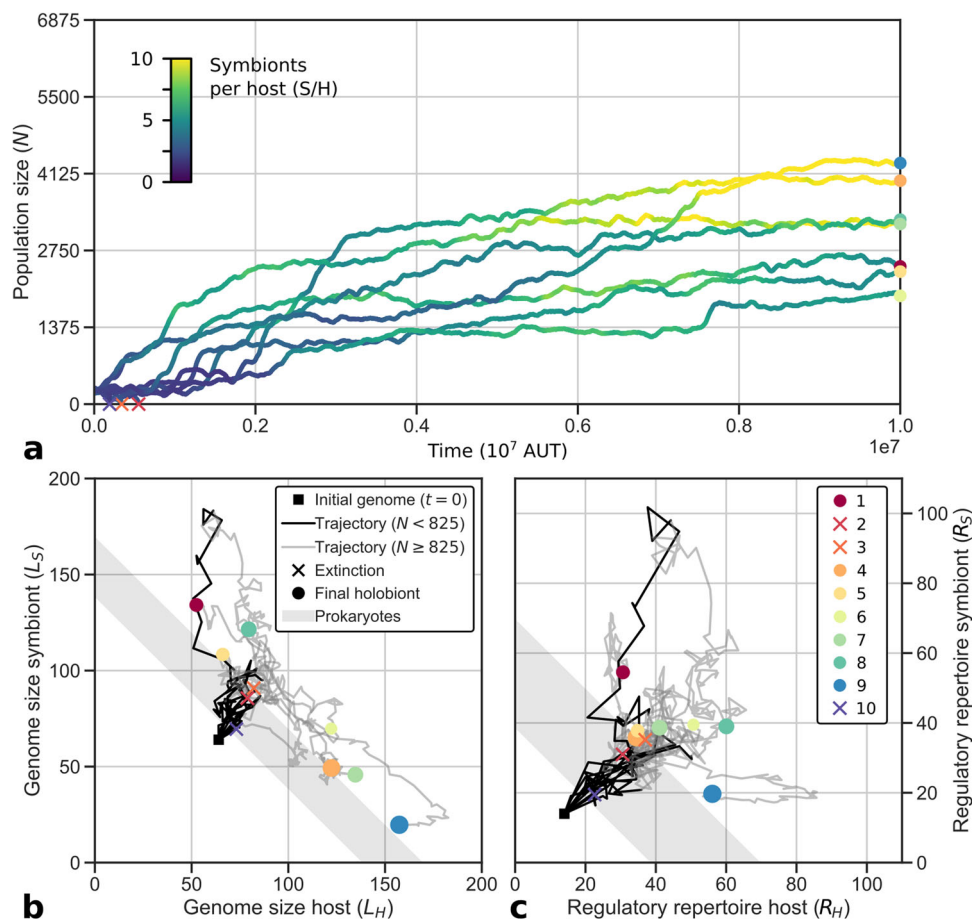

Fig. 2 Primitive holobionts adapt to nutrient gradient. a Population size increases, coinciding with an increase in symbiont numbers. **b** Genomes expand early, and later genome size symmetry breaks between host and symbiont. **c** Regulatory repertoires expand further than free-living cells for all hosts and symbionts that survive. The maximum population size is given by the grid size: $25 \times 275 = 6875$. Adaptation generally occurs before $t = 10^7$ (see Supplementary Fig. S2 for trajectories until $t = 2 \cdot 10^7$). In panels (**b**) and (**c**), the size of the markers is scaled by the final population size.
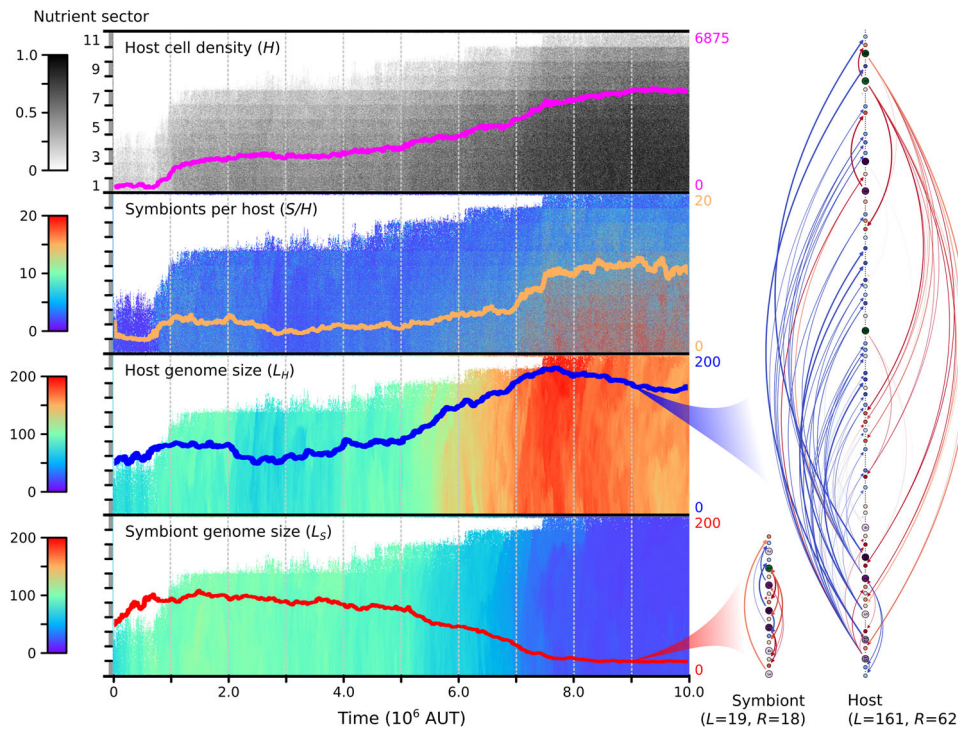
Fig. 3 Evolutionary dynamics of the most successful replicate with primitive FECA, P9. Each panel shows a different variable in spacetime. The overlaid lines represent summaries across the entire gradient (for the top panel, the overlaid line shows total population size rather than local cell density, in correspondence with Fig. 2a). On the right, genomes of the host and symbiont in the ancestor at $t = 9 \cdot 10^6$ are depicted from top to bottom, along with the regulatory interactions (L denotes total genome size, R regulatory repertoire size). For an example of a different evolutionary trajectory, see Supplementary Fig. S3.

the first substantial adaptations in the seven replicates that survive (e.g., the adaptation seen in P9 around $t = 10^6$ in Fig. 3). These adaptations lead to an increase in population size, after which selection intensifies and genome expansion slows down.

At the end of the experiment, the regulatory repertoires of hosts and symbionts have surpassed what we observed for free-living prokaryotes (the gray band in Fig. 2c; cf. Von der Dunk et al.[29]). Endosymbiosis imposes constraints on host and symbiont that are absent in free-living cells and which weaken selection for genomic streamlining. Thus, the endosymbiotic condition contributes to genome expansion and potentially explains some of the complexity that emerged during eukaryogenesis.

Concurrent expansion and shrinkage of host and symbiont genomes. On long timescales, the genome sizes of the host and symbiont diverge: four replicates evolve toward large host and small symbiont genomes, and three replicates evolve toward large symbiont and small host genomes (Fig. 2b). The first of these two scenarios, which resembles the outcome of eukaryogenesis, is also more common in the evolution experiment starting with complex FECA (C1–12). Specifically, the replicates that are initialized with identical host and symbiont (C1–4) always evolve to have a somewhat larger host than symbiont genome. Moreover, in both evolution experiments (P1–10 and C1–12), the replicates that evolved a larger host than the symbiont genome reached a larger population size (see Supplementary Fig. S4). In P9, the dramatic symmetry breaking in genome size that takes place from around $t = 4 \cdot 10^6$ coincides with an increase in population size, revealing adaptation until $t = 8 \cdot 10^6$. Thus, successful endosymbiosis leads to a complex host and a simple symbiont, similar to the outcome of eukaryogenesis.

Evolution of symbiont numbers. Symbiont numbers constitute an important degree of freedom of our model that is subject to enormous variation arising at different levels. For an individual holobiont, there is already an associated probability distribution of symbiont numbers due to inherent stochasticity in regulatory dynamics and due to changes in host and symbiont cell-cycle behavior as a result of variation in nutrient conditions. For the entire population, mutations create differences in these probability distributions between individual holobionts, which allows them to evolve over time. Below, we describe how variation in symbiont numbers across time and space has contributed to holobiont adaptation, starting at the population level, then zooming in to individual holobionts, and finally focusing on the details of host and symbiont cell-cycle behavior within a single holobiont.

During the evolution experiment initialized with primitive FECA, the average number of symbionts per host in the population increases from 2.5–2.8 to 4.3–10.2 across replicates (Fig. 2a). The increase in symbiont numbers is directly linked to holobiont adaptation. In P9, for instance, both adaptive phases that we identified before are accompanied by marked increases in symbiont numbers (Fig. 3). More generally, final population size correlates well with symbiont number across replicates ($r = 0.68$; Supplementary Fig. S5). Thus, symbiont numbers play an important role in adaptation toward a successful endosymbiotic lifestyle. On top of the increased average symbiont number at the end of replicate P9, we find that symbiont numbers have diversified across the nutrient gradient (this is the case for about half of the replicates; other populations evolved low symbiont numbers and little diversity across the gradient, see, e.g., Supplementary Fig. S3).
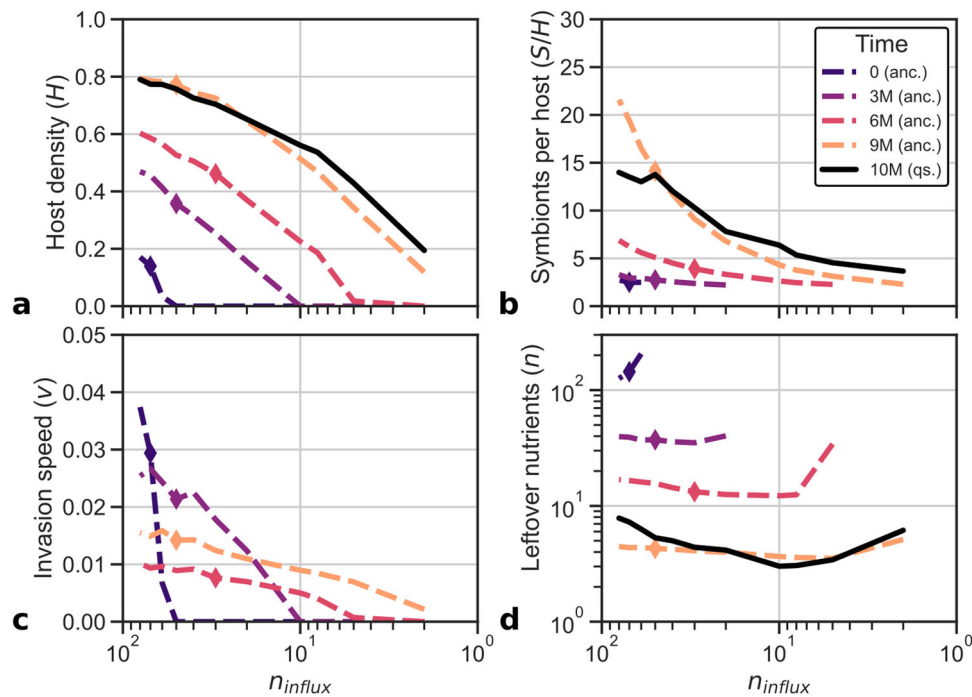
Fig. 4 Emergence of an r- and K-strategy along the ancestry of P9. Over time, cell density (**a**) and symbiont number (**b**) have increased, but growth measured as invasion speed (**c**) and nutrient availability (**d**) have decreased. The final population resembles the most recent ancestor, but strains at high and low nutrient influxes have adapted to those specific environments. Diamonds depict the conditions in which the ancestor lived.

Holobionts adapt through r- or K-strategy. To understand the observed evolutionary patterns in genome size and symbiont number, we now study adaptive behavior at the level of individual holobionts, which will subsequently be broken down further into the underlying host and symbiont behaviors. First, to discern the strategies by which holobionts have adapted, we studied clonal population dynamics of four ancestors extracted at different time points along the evolutionary trajectory of P9 under the nutrient conditions of each sector of the gradient (see "Characterizing holobionts" in "Methods"; Fig. 4). As expected, these ancestors broadly recapitulate the adaptation that was observed in the full population, i.e., increases in host cell density and symbiont numbers (Fig 4a, b). In addition, holobiont profiling reveals two alternative strategies (Fig. 4). The early ancestors at $t = 0$ and $t = 3 \cdot 10^6$ maintain very few symbionts, yielding unstable holobionts and low cell density. As a consequence, nutrients are not depleted much, and holobionts grow (i.e., invade) fast in those conditions where they are viable ($n_{influx} < 50$ and $n_{influx} < 10$, respectively). Conversely, the more recent ancestors at $t = 6 \cdot 10^6$ and $t = 9 \cdot 10^6$ maintain many symbionts, yielding stable holobionts and high cell density. Yet nutrients are depleted to low levels, and holobionts grow slowly. Thus, through the evolution of symbiont numbers, holobionts have access to an r-strategy which optimizes growth (r) and a K-strategy which optimizes carrying capacity (K). Early in evolution, when empty space and nutrients are abundant, selection for fast invasion favors the r-strategy, whereas later, as empty space and nutrients become limited due to improved cell-cycle behavior, the K-strategy is more advantageous. Nevertheless, some populations settle on an r-strategy (e.g., P8, C10 and C12; see also "Genome size evolution directs holobiont adaptation"), highlighting the role of historical contingency and chance processes in evolution.

Individual- and population-level regulation of symbiont numbers. Surprisingly, the previously observed variation in symbiont numbers across the nutrient gradient at the end of replicate P9 (Fig. 3) is for a large part explained by individual regulation of symbiont numbers with nutrient conditions as observed in the ancestor profiles (Fig. 4b). When the holobiont encounters poor nutrient conditions, it maintains few symbionts, and when it encounters rich nutrient conditions, it maintains many symbionts. This emergent control ensures that there are always sufficient nutrients available for the host to finish genome replication (as seen in Fig. 4c) and that resources are only diverted to symbionts to increase holobiont stability when conditions allow it. Thus, the holobiont has evolved a mechanism to successfully navigate the large fluctuations in nutrient levels resulting from local variations in host cell density and symbiont numbers.

The full quasi-species at $t = 10^7$ diverges only slightly from the individual symbiont control observed in the ancestor at $t = 9 \cdot 10^6$. In environments that are poorer than the native environment of the ancestor ($n_{influx} = 50$), strains of the quasi-species have adapted by boosting average symbiont numbers relative to the ancestor (e.g., from 2.3 to 3.7 at $n_{influx} = 2$), enhancing holobiont stability and carrying capacity, and reducing growth. Conversely, in richer environments, strains have adapted by reducing symbiont numbers relative to the ancestor (e.g., from 21.6 to 14.0 at $n_{influx} = 80$), promoting growth while barely reducing carrying capacity—since stability is already almost guaranteed with more than ten symbionts per host.

Emergent symbiont control through implicit cell-cycle coordination. To understand how the adaptations to endosymbiosis at the holobiont level come about—i.e., r- and K-strategies and symbiont control—we analyze the cell-cycle behavior of host and symbiont independently (see "Tracking cell-cycle behavior" in "Methods"). By plotting their growth curves as a function of nutrient condition (see "Calculating growth rate" in "Methods"), we obtain a phase diagram for the holobiont (Fig. 5, left panels). For any particular nutrient condition, the relative positions of the growth curves reveal whether the symbiont number will increase
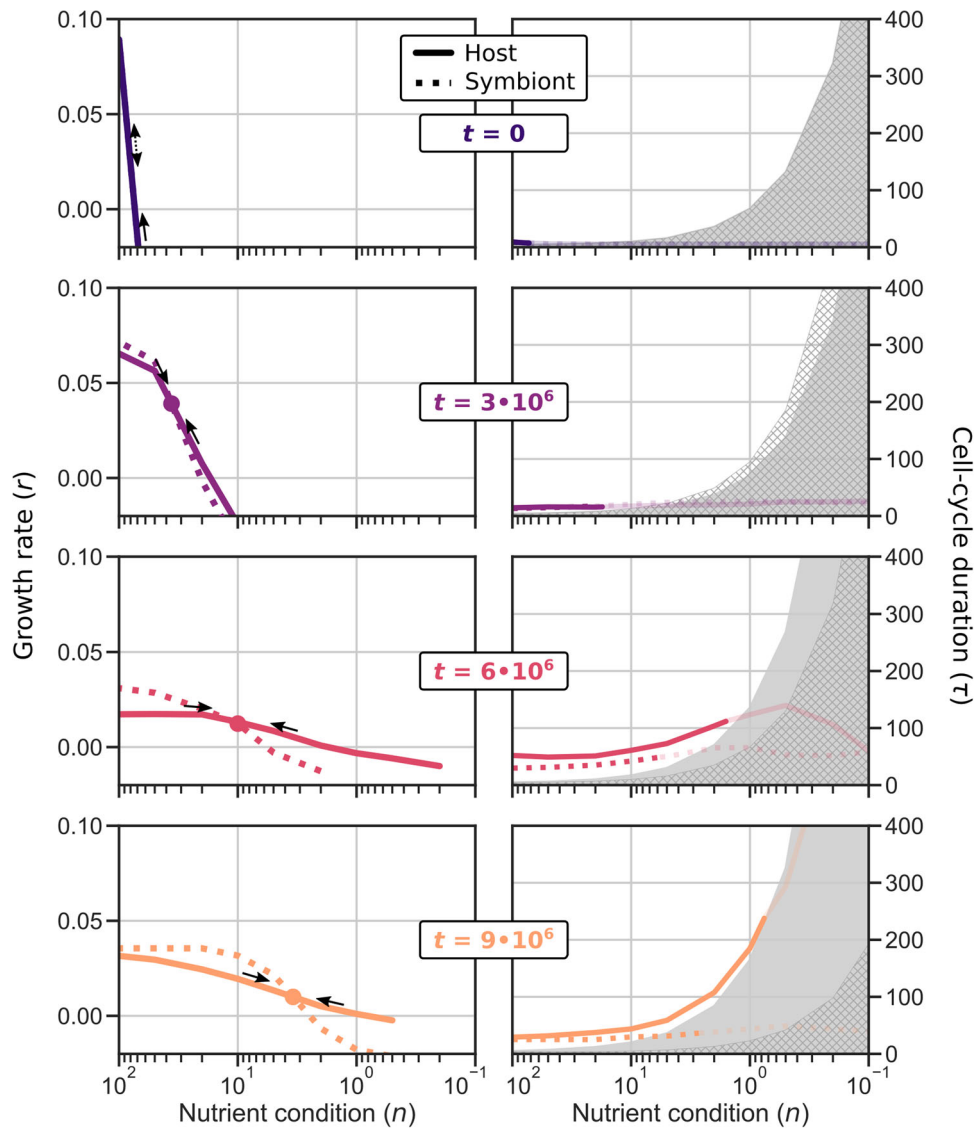
Fig. 5 Evolution of host–symbiont cell-cycle coordination. Host–symbiont cell-cycle coordination evolves, first through differential cell-cycle timing, later through niche differentiation whereby the host becomes a generalist and the symbiont a specialist. Each row shows the growth rates (left panels) and cell-cycle durations (right panels) of an ancestor of the P9 replicate. In contrast to Fig. 4, the x-axis here depicts fixed nutrient conditions where host and symbiont cell-cycle dynamics are assessed independently. The gray areas in the cell-cycle duration panels mark the minimum possible duration of a successful cell cycle (with enough time for replication) given the genome size of the host (filled) and symbiont (hatch). The cell-cycle duration is only drawn in bold for the range where $\rho > 0.5$, which roughly coincides with $r > 0$.

($r_S > r_H$) or decrease ($r_S < r_H$) with time. The change in symbiont number translates to a shift on the x-axis: an increase in symbiont number corresponds to a decrease in nutrient abundance (shift to the right), and a decrease in symbiont number corresponds to an increase in nutrient abundance (shift to the left).

The phase diagrams depicting the growth of ancestral hosts and symbionts reveal how holobionts achieved symbiont control during evolution. Initially, symbiont dynamics are unstable: even with equal host and symbiont growth rates across nutrient conditions, stochasticity causes symbiont number to drift up or down, resulting in holobiont death due to symbiont loss or host starvation (i.e., $r_H < 0$ for $n \lesssim 70$ for the initial holobiont). Later holobionts ($t \geq 3 \cdot 10^6$) managed to stabilize symbiont dynamics through the differentiation of host and symbiont growth. Perturbations from the stable equilibrium (i.e., the intersection of r_H and r_S) are compensated by subsequent changes in relative symbiont and host growth rate that push the system back to this equilibrium. The evolved stable equilibrium also explains

previously observed holobiont behavior: host–symbiont growth dynamics equilibrate at the same nutrient level regardless of nutrient influx (Fig. 4d), which means that with greater nutrient influx, more symbionts can be sustained before nutrients are depleted to that equilibrium level (Fig. 4b).

Given that we understand how holobionts control symbiont numbers, we can clearly observe how cells evolved higher symbiont numbers over time (Fig. 5). Between $t = 3 \cdot 10^6$ and $t = 9 \cdot 10^6$, the position of the stable growth point moves from high nutrient abundance (few symbionts) and high growth rate (r-strategy) to low nutrient abundance (many symbionts) and low growth rate (K-strategy). Underlying this transition from r- to K-strategy are changes in the cell-cycle behavior. First, the host growth curve is flattened when the host has evolved a slower cell cycle and increased survival in poor conditions (e.g., $t = 6 \cdot 10^6$). Later ($t = 9 \cdot 10^6$), the host and symbiont diverge dramatically in genome size and differentiate into a generalist and specialist, respectively, referring to their ability (and inability) to tune cell-
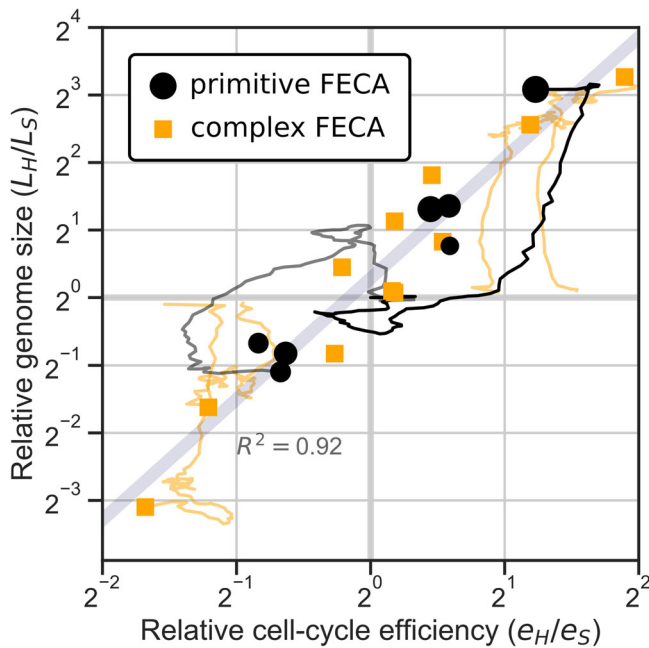
Fig. 6 Asymmetry in genome size correlates strongly with asymmetry in cell-cycle efficiency between host and symbiont. Along the ancestral lineage, genome size and cell-cycle efficiency are measured (see "Calculating cell-cycle efficiency" in "Methods"). Trajectories are shown for the experiments where the largest genome size asymmetry evolved (P1, P9 and C9–12); the trajectory of P9 is in bold. One technical replicate is shown for each of the replicates initialized with complex FECA. For the experiment initialized with primitive FECA, the size of the markers is scaled by the final population size.

cycle behavior to nutrient conditions[29]. The invention of generalism allows the holobiont to maintain more symbionts in equilibrium, increasing stability while also promoting faster growth when provided with more nutrients at the invasion front (i.e., see Fig. 4c).

Cell-cycle coordination and genome size evolution. The characterization of cell-cycle coordination between host and symbiont allows us to revisit emergent genome size asymmetry, i.e., the appearance of complex hosts with simple symbionts or vice versa (Fig. 2b). For this, we turn our focus to the evolution replicates starting with a complex FECA (C1–12). It turns out that when a prokaryote with a more efficient cell cycle (i.e., R8 or R9)—quantified as $e = \frac{\tau_{min}}{\tau}$ (see "Calculating cell-cycle efficiency" in "Methods")—is matched with a prokaryote with a less efficient cell cycle (e.g., R2 or R3), the former evolves a large genome and the latter a small genome, irrespective of which prokaryote is the host and which the symbiont. Across both evolution experiments, asymmetry in cell-cycle efficiency explains 92% of the variation in genome size asymmetry (Fig. 6).

The strong correlation between cell-cycle efficiency and genome size in host–symbiont pairs is the result of selection for cell-cycle coordination. To balance growth between the host and symbiont, selection forces the more efficient partner to slow down its cell cycle to the speed of the less efficient partner. The added cell-cycle time is utilized by the more efficient partner to replicate more household genes, relieving the less efficient partner of this burden. Thus, genome size asymmetry compensates for differences in cell-cycle efficiency between the host and symbiont, allowing for the faster overall growth of the holobiont. This mechanism applies when asymmetry in cell-cycle efficiency is ingrained in the holobiont from the start, as in the evolution

experiment with complex FECA, but also when the asymmetry arises spontaneously, as in the evolution replicates with primitive FECA where, by chance, either the host or symbiont acquires more efficient cell-cycle regulation.

A similar adaptive mechanism explains why hosts evolve larger genomes than symbionts when they start out identical (C1–4). Hosts generally evolve slower cell cycles than symbionts in order to stabilize symbiont dynamics (i.e., to obtain the stable dynamic equilibrium, Fig. 5), so they can spend more time replicating household genes and partly relieve the symbiont.

As previously mentioned, non-adaptive forces play a large role in genome expansion and genome size early in evolution. Yet the final asymmetry in genome size is dominated by adaptive forces compensating differences in cell-cycle efficiency between host and symbiont (Fig. 6). When household genes cannot be shared between host and symbiont, a small but consistent asymmetry remains (Supplementary Fig. S6), indicating that a host or symbiont with greater replication capacity due to a slower cell cycle experiences more genome expansion than its partner even when there is no compensatory advantage by concurrent streamlining of that partner genome. In short, large genomes are less costly when a slow cell cycle is favored by other constraints. However, genome size evolution is not only a by-product of cell-cycle coordination but also directs holobiont adaptation toward an r- or K-strategy, as explained in detail in "Genome size evolution directs holobiont adaptation" in Supplementary Note 1.

Discussion

Endosymbiosis begets complexity. We forced different combinations of cells with their own genome and gene regulatory network into an obligate endosymbiotic relationship. In the vast majority of our experimental replicates, holobionts evolved implicit cell-cycle coordination through differential growth behavior on the same resource. This shows that no complex cellular traits are in principle required to control symbionts other than the ability of the host and symbiont to adapt to different nutrient conditions, i.e., no intracellular communication via protein targeting is required and no advanced cell-cycle control by the host.

Nevertheless, not all cells become equally successful hosts or symbionts. Pre-evolved cells, which already performed complex cell-cycle regulation, adapted much faster to endosymbiosis and reached greater population sizes than primitive cells. As in free-living cells, the efficiency of cell-cycle regulation is an important factor in successful holobiont adaptation. In addition, generalist behavior is preferred for hosts, allowing them to deal with both high and low nutrient supplies resulting from fluctuations in symbiont numbers. For symbionts, relatively simple specialist behavior, as observed in P9, appears sufficient for successful holobionts. In nature, endosymbionts experience a relatively constant environment inside their host, which has been hypothesized to drive genomic streamlining[26,27]. In our model, nutrient homeostasis is accomplished through indirect control of symbiont numbers and by adaptation of the host to the remaining fluctuations, which also allows the symbiont to specialize on specific nutrient conditions and streamline its genome.

Genome size evolution. During eukaryogenesis, genome size increased immensely[16]. In our model, the total genome size of the holobiont also increased with respect to free-living cells, where genome expansion was limited but necessary for functional adaptation[29]. Moreover, under the high mutation rates of the prokaryote-like regime, holobionts experienced even more genome expansion and went extinct frequently, indicating that the

endosymbiotic condition weakens selection for genomic stream-lining. These outcomes are in line with the non-adaptive account of genome expansion proposed by Lynch[21]. When an organism has more tasks to perform than replicating as fast as possible, the selection of small genomes to speed up replication is weakened.

Another striking pattern in the outcome of eukaryogenesis is the asymmetry in genome size between the host (nuclear) and symbiont (mitochondrial) genomes[32]. Most of the proposed drivers of this asymmetry—such as increased substitution rate in the symbiont due to lack of recombination and ROS production[18,19] or advantageous host regulatory control[17]—are not incorporated into our model, yet asymmetry still evolves. In our model, unequal replication efficiency drives genome size asymmetry. Apart from the outcome of eukaryogenesis, i.e., large host and small symbiont genome, this mechanism also yields the opposite outcome, i.e., small host and large symbiont genome, which has not been observed in any endosymbiosis in nature. Interestingly, this opposite outcome could be a model for the first step in the Syntrophy hypothesis[9], whereby a symbiotic archaeon inside a delta-proteobacterial host eventually gave rise to a large nuclear genome. Note, however, that the outcome with a large host and small symbiont genome is favored because on short timescales, cell-cycle coordination selects for a slow host, yielding asymmetric constraints on genome size, and on long timescales, holobionts with large host and small symbiont genomes evolve the highest carrying capacity (K-strategy) outcompeting all other holobionts.

Since all extant endosymbiotic relationships involve a host that is already complex, host control over symbiont numbers has not yet been addressed as a potential hurdle for eukaryogenesis. Yet, for eukaryogenesis, we do not know whether targeted protein transport through a complex endomembrane system was already possible. If host control over symbiont numbers evolved later during eukaryogenesis, the number of symbionts per holobiont would initially have been a key emergent property of host and symbiont growth dynamics. As we show in our relatively simple model, where symbiont numbers impact holobiont behavior and interact with the environment, multiple holobiont strategies are possible. The evolved r- and K-strategies are tightly linked to genome size evolution (such a link was also pointed out in Cavalier-Smith[33]). In the K-strategy, symbionts exist in large populations and experience strong selection for genomic streamlining and fast growth. In contrast, in the r-strategy, symbionts evolve large genomes and exist in small populations, which can lead to a collapse of the population much later on (e.g., in C12 around $t = 8 \cdot 10^6$, far outside the range shown in Supplementary Fig. S1).

Transition in individuality

We have studied eukaryogenesis as a prototype of a major transition in evolution[8], in which self-sufficient entities give up their self-sufficiency and revise the balance of selection. In contrast to previous models of such transitions[34,35], cell division of the host is not triggered by reaching a predefined number of lower-level entities but is effectuated by the host's own cell cycle. The consequential variation in symbiont number bestows evolution with an additional degree of freedom. Symbiont number not only shapes the holobiont phenotype, i.e., stability and growth rate but also the strength of selection at the symbiont level relative to the holobiont level. It is well-known that the selection balance between symbiont and holobiont levels also depends on their timescales[11]. Here the timescales are defined by the duration of symbiont and host cell cycles. Thus, in line with theoretical work on the RNA world[34,35], selection is itself tuned by evolution in various ways. Cell biology, ecology and evolution are constantly interacting to shape the characteristics of living organisms in response to their environment. Our model shows how these interactions could explain the complexification of the eukaryotic cell.

Methods

Characterizing holobionts. To study how cell cycles of the host and symbiont adapted to endosymbiosis, we profiled holobionts in the ancestral lineage of P9. For this, we performed small experiments with clonal populations of a single holobiont type (without mutation) on every sector of the gradient. First, propagation across empty space is measured, followed by bulk measures of cell density, symbionts per host and leftover nutrients. We also measured three of the four observables directly in the gradient at the end of replicate P9 ($t = 10^7$) to see if and how diversity in the quasi-species has improved adaptation to the gradient relative to the ancestral holobiont at $t = 9 \cdot 10^6$.

Tracking cell-cycle behavior. We performed experiments where individual hosts and symbionts execute their cell-cycle behavior for 10^4 AUT under different fixed nutrient conditions as described in ref. [29]. During the experiment, we tracked the average cell-cycle duration τ, the fraction of successful cell cycles ρ, and calculated the instantaneous growth rate as $r = \frac{\rho}{\tau}(1 - \frac{1}{R_0})$ where $R_0 = \frac{\rho}{\delta\tau + 1 - \rho}$ (based on an ODE description of the model, see "Calculating growth rate" and Von der Dunk et al.[29]).

Calculating growth rate. In the Supplementary Material of our previous study[29], we presented a simple ODE model that describes phenomenologically a cell population (N) executing a fixed cell cycle under constant nutrient conditions.

$$\frac{dN}{dt} = \frac{\rho}{\tau}N(1 - N) - \frac{1 - \rho}{\tau}N - \delta N \quad (1)$$

From the cell-cycle parameters τ and ρ (which are measured in single-cell simulation runs), we derived the reproduction number R_0 as:

$$R_0 = \frac{\rho}{\delta\tau + 1 - \rho} \quad (2)$$

Here we derive the instantaneous growth rate r allowing us to compare the effective growth of the symbiont relative to the host across nutrient conditions. Similar to R_0, the instantaneous growth rate is defined in optimal conditions (i.e., $N \to 0$) as the per capita growth rate:

$$g(N) = \frac{\rho}{\tau}(1 - N) - \frac{1 - \rho}{\tau} - \delta \quad (3)$$

$$r = g(N \to 0) = \frac{2\rho - 1 - \delta\tau}{\tau} \quad (4)$$

$$r = \frac{\rho}{\tau}\left(1 - \frac{1}{R_0}\right) \quad (5)$$

Calculating cell-cycle efficiency. The fastest possible cell cycle for a given fixed nutrient abundance n and genome size L is $\tau_{min} = 3 + L/n$. The duration of an actual cell cycle relative to this fastest possible cell cycle gives the efficiency $e = \frac{\tau_{min}}{\tau} = \frac{3 + L/n}{\tau}$. We average efficiencies obtained under conditions $n \in \{100, 50, 20, 10, 5, 2, 1\}$ from the single-cell experiments (see "Tracking cell-cycle behavior") to arrive at a single value for an individual.

Statistics and reproducibility. Details of the evolution experiments, including population sizes and number of replicates, are provided in the main text and in "Methods". We report on correlations at several occasions, all of which are Pearson correlations: population size × symbiont number ($r = 0.68$, $R^2 = 0.46$, $p = 0.0019$, $N = 18$; Supplementary Fig. S5 and main text), relative genome size × relative cell-cycle efficiency ($R^2 = 0.92$; Fig. 6 and main text), population size × genome size host ($R^2 = 0.23$, $p = 0.046$, $N = 18$; Supplementary Fig. S4), relative genome size × relative cell-cycle efficiency without sharing of household genes ($R^2 = 0.76$, $p = 4.43 \cdot 10^{-6}$, $N = 17$).

Data availability

Data were generated from a custom-built computational model written in C++ (see "Code availability"). Key genome data are available at https://doi.org/10.5061/dryad.tx95x6b2v; all other data are available upon request.

Code availability

Annotated code is available at https://github.com/samvonderdunk/Eukaryotes.

References

1. Sagan, L. On the origin of mitosing cells. *J. Theor. Biol.* **14**, 225–274 (1967).
2. Kurland, C., Collins, L. & Penny, D. Genomics and the irreducible nature of eukaryote cells. *Science* **312**, 1011–1014 (2006).
3. Martin, W. & Koonin, E. V. Introns and the origin of nucleus–cytosol compartmentalization. *Nature* **440**, 41–45 (2006).
4. Stairs, C. W. & Ettema, T. J. The archaeal roots of the eukaryotic dynamic actin cytoskeleton. *Curr. Biol.* **30**, R521–R526 (2020).
5. Lane, N. & Martin, W. The energetics of genome complexity. *Nature* **467**, 929–934 (2010).
6. Raval, P. K., Garg, S. G. & Gould, S. B. Endosymbiotic selective pressure at the origin of eukaryotic cell biology. *Elife* **11**, e81033 (2022).
7. Schavemaker, P. E. & Muñoz-Gómez, S. A. The role of mitochondrial energetics in the origin and diversification of eukaryotes. *Nat. Ecol. Evol.* **6**, 1307–1317 (2022).
8. Szathmáry, E. & Maynard Smith, J. The major evolutionary transitions. *Nature* **374**, 227–232 (1995).
9. López-García, P. & Moreira, D. The Syntrophy hypothesis for the origin of eukaryotes revisited. *Nat. Microbiol.* **5**, 655–667 (2020).
10. Roughgarden, J., Gilbert, S. F., Rosenberg, E., Zilber-Rosenberg, I. & Lloyd, E. A. Holobionts as units of selection and a model of their population dynamics and evolution. *Biol. Theory* **13**, 44–65 (2018).
11. Van Vliet, S. & Doebeli, M. The role of multilevel selection in host microbiome evolution. *Proc. Natl Acad. Sci. USA* **116**, 20591–20597 (2019).
12. Colizzi, E. S., Hogeweg, P. & Vroomans, R. M. Modelling the evolution of novelty: a review. *Essays Biochem.* **66**, 727–735 (2022).
13. Colizzi, E. S. & Hogeweg, P. Evolution of functional diversification within quasispecies. *Genome Biol. Evol.* **6**, 1990–2007 (2014).
14. Colizzi, E. S., Vroomans, R. M. & Merks, R. M. Evolution of multicellularity by collective integration of spatial information. *Elife* **9**, e56349 (2020).
15. Zachar, I. & Szathmáry, E. Breath-giving cooperation: critical review of origin of mitochondria hypotheses: major unanswered questions point to the importance of early ecology. *Biol. Direct* **12**, 1–26 (2017).
16. Vosseberg, J. et al. Timing the origin of eukaryotic cellular complexity with ancient duplications. *Nat. Ecol. Evol.* **5**, 92–100 (2021).
17. Nowack, E. C. *Paulinella chromatophora*—rethinking the transition from endosymbiont to organelle. *Acta Soc. Bot. Pol.* **83**, 387–397 (2014).
18. Allen, J. F. & Raven, J. A. Free-radical-induced mutation vs redox regulation: costs and benefits of genes in organelles. *J. Mol. Evol.* **42**, 482–492 (1996).
19. Martin, W. & Herrmann, R. G. Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol.* **118**, 9–17 (1998).
20. Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**, 123–135 (2004).
21. Lynch, M. & Conery, J. S. The origins of genome complexity. *Science* **302**, 1401–1404 (2003).
22. Stoltzfus, A. On the possibility of constructive neutral evolution. *J. Mol. Evol.* **49**, 169–181 (1999).
23. Finnigan, G. C., Hanson-Smith, V., Stevens, T. H. & Thornton, J. W. Evolution of increased complexity in a molecular machine. *Nature* **481**, 360–364 (2012).
24. Ba, A. N. N. et al. Parallel reorganization of protein function in the spindle checkpoint pathway through evolutionary paths in the fitness landscape that appear neutral in laboratory experiments. *PLoS Genet.* **13**, e1006735 (2017).
25. Nowack, E. C. et al. Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of *Paulinella chromatophora*. *Proc. Natl Acad. Sci. USA* **113**, 12214–12219 (2016).
26. Gabr, A., Stephens, T. G. & Bhattacharya, D. Loss of key endosymbiont genes may facilitate early host control of the chromatophore in *Paulinella*. *iScience* **25**, 104974 (2022).
27. McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26 (2012).
28. Quiñones-Valles, C., Sánchez-Osorio, I. & Martínez-Antonio, A. Dynamical modeling of the cell cycle and cell fate emergence in *Caulobacter crescentus*. *PLoS ONE* **9**, e111116 (2014).
29. Von der Dunk, S. H., Snel, B. & Hogeweg, P. Evolution of complex regulation for cell-cycle control. *Genome Biol. Evol.* **14**, evac056 (2022).
30. Guerrero, R., Margulis, L. & Berlanga, M. Symbiogenesis: the holobiont as a unit of evolution. *Int. Microbiol.* **16**, 133–143 (2013).
31. Gould, S. J. & Eldredge, N. Punctuated equilibria: an alternative to phyletic gradualism. *Models Paleobiol.* **1972**, 82–115 (1972).
32. Roger, A. J., Muñoz-Gómez, S. A. & Kamikawa, R. The origin and diversification of mitochondria. *Curr. Biol.* **27**, R1177–R1192 (2017).
33. Cavalier-Smith, T. r- and K-tactics in the evolution of protist developmental systems: cell and genome size, phenotype diversifying selection, and cell cycle patterns. *Biosystems* **12**, 43–59 (1980).
34. Takeuchi, N. & Hogeweg, P. Multilevel selection in models of prebiotic evolution II: a direct comparison of compartmentalization and spatial self-organization. *PLoS Comput. Biol.* **5**, e1000542 (2009).
35. Takeuchi, N., Hogeweg, P. & Koonin, E. V. On the origin of DNA genomes: evolution of the division of labor between template and catalyst in model replicator systems. *PLoS Comput. Biol.* **7**, e1002024 (2011).

Author contributions
S.D., P.H. and B.S. designed the research. S.D. performed the experiments and analysis and wrote the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-023-05153-x.

Correspondence and requests for materials should be addressed to Samuel H. A. von der Dunk.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary handling editor: Christina Karlsson-Rosenthal.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.