Encounter networks from collective mitochondrial dynamics support the emergence of effective mtDNA genomes in plant cells

Konstantinos Giannakis[1], Joanna M. Chustecki[2], Iain G. Johnston[1,3,*]

[1] Department of Mathematics, University of Bergen, Bergen, Norway

[2] School of Biosciences, University of Birmingham, Birmingham, UK

[3] Computational Biology Unit, University of Bergen, Bergen, Norway

[*] correspondence to `iain.johnston@uib.no`

Abstract

Mitochondria in plant cells form strikingly dynamic populations of largely individual organelles. Each mitochondrion contains on average less than a full copy of the mitochondrial DNA (mtDNA) genome. Here, we asked whether mitochondrial dynamics may allow individual mitochondria to 'collect' a full copy of the mtDNA genome over time, by facilitating exchange between individuals. Akin to trade on a social network, exchange of mtDNA fragments across organelles may lead to the emergence of full 'effective' genomes in individuals over time. We characterise the collective dynamics of mitochondria in *Arabidopsis thaliana* hypocotyl cells using a recent approach combining single-cell timelapse microscopy, video analysis, and network science. We then use a quantitative model to predict the capacity for the sharing and accumulation of genetic information through the networks of encounters between mitochondria. We find that biological encounter networks are strikingly well predisposed to support the collection of full genomes over time, outperforming a range of other networks generated from theory and simulation. Using results from the coupon collector's problem, we show that the upper tail of the degree distribution is a key determinant of an encounter network's performance at this task and discuss how features of mitochondrial dynamics observed in biology facilitate the emergence of full effective genomes.

Introduction

Mitochondria are vital bioenergetic organelles, present in the vast majority of eukaryotic cells. Across and within eukaryotic organisms, mitochondria display a diverse variety of forms and dynamics. In plant cells, mitochondria largely exist as discrete, independent organelles. Unlike metazoan and fungal mitochondria, they rarely form large physical networks (with some exceptions [Seguí-Simarro and Staehelin, 2009]). Individual plant mitochondria are highly dynamic, moving rapidly through the cell both along the cytoskeleton and diffusively [Logan and Leaver, 2000, Logan, 2006].

This physical population has a coupled genetic structure. Plant mitochondria do not typically contain full copies of the mtDNA genome [Preuten et al., 2010, Takanashi et al., 2006, Johnston, 2019]. Instead, many mitochondria either contain mtDNA 'subgenomic' molecules – encoding a reduced subset of mtDNA genes – or no mtDNA at all. The question arises: how do plant mitochondria maintain their protein complements, without a complete local genome from which to express new proteins?

One possibility [Arimura et al., 2004, Logan, 2006, Takanashi et al., 2006, Arimura, 2018] is that exchanges of mtDNA subsets between individuals can, over time, lead to the emergence of full 'effective' mtDNA genomes in individuals over time. For example, picture a genome which can be partitioned into two regions, A and B. One mitochondrion initially possesses a subgenomic molecule containing only region A of the genome. Another initially possesses only region B. Each expresses the genes contained in its subgenomic region. Then the two mitochondria physically meet and exchange their subgenomic molecules. The first mitochondrion can now express genes from B, and vice versa. Indeed, within the dynamic cellular population of mitochondria, transient colocalisations occur, resembling 'kiss-and-run' events in bacterial populations [Liu et al., 2009, Logan, 2010, El Zawily et al., 2014, Chustecki et al., 2021]. Some of these colocalisations result in transient fusion between two mitochondria. When fusion occurs, mitochondria can exchange genetic and protein material: indeed, mixing occurs through the entire cellular population on a timescale of hours [Arimura et al., 2004].

Recent work has characterised the 'encounter networks' between mitochondria in plant cells, describing which mitochondria encounter which others over time [Chustecki et al., 2021]. Here, mitochondria are nodes, with two nodes linked by an edge if the corresponding mitochondria have been recorded within a threshold distance. Chustecki *et al.* showed that these encounter networks have structures which have the potential to facilitate efficient exchange of content, while also allowing mitochondria to spread evenly through the cell [Chustecki et al., 2021]. Hence, mitochondrial dynamics have the potential to resolve a tension between competing cell priorities: even spacing of mitochondria (with metabolic and energetic advantages) and colocalisation of mitochondria (for beneficial exchange of contents). These principles support the developing cell biological perspective of inter-organelle interactions [Valm et al., 2017, Cohen et al., 2018, Picard and Sandi, 2021].

Such functional encounters are an example of emergence, where the behaviour of a collective of individuals is different from the sum of individual behaviours. There are two coupled instances of emergent behaviour in our system – physical and genetic. First, the encounter network of mitochondria emerges from their underlying physical dynamics in the cell [Williams and George, 2019]. Second, through genetic exchanges within this encounter network, an 'effective genome' for each mitochondrion may emerge. That is, over time, each mitochondrion will be exposed to a growing set of genetic information. We hypothesised that the exchange efficiency of encounter networks could allow a mechanism for plant mitochondria to address their maintenance problem. Specifically, if mitochondria can efficiently exchange genetic information, then the effective genome, to which each mitochondrion is exposed over time, may eventually grow to include the full set of genes in the full genome. To investigate this hypothesis, we proceed by using network science and quantitative modelling of exchange processes to investigate the genetic behaviours that these encounter networks could potentially support.

Results

The emergence of effective genomes on *Arabidopsis* encounter networks as a network science problem

We first sought to understand the process by which effective genomes could potentially emerge from dynamic interchange of subgenomic molecules in plant cells, using encounter networks characterised from hypocotyl cells in 7-day *Arabidopsis* seedlings (see Methods). In previous work, we established an experimental and computation pipeline to characterise the 'social' encounter networks of mitochondria [Chustecki et al., 2021]. Here, nodes represent mitochondria, and an edge between two nodes means that those two mitochondria have colocalised within a physical threshold distance at at least one timepoint during the experiment (Fig. 1). Example networks from mitochondrial dynamics in *Arabidopsis* hypocotyl are shown in Fig. 1. We will use these, and other experimentally-characterized encounter networks, in the subsequent analysis. Results from independent single cells were generally very similar (Fig. S8).

We proceed by phrasing the core biological question – the collection of genetic information by mitochondria
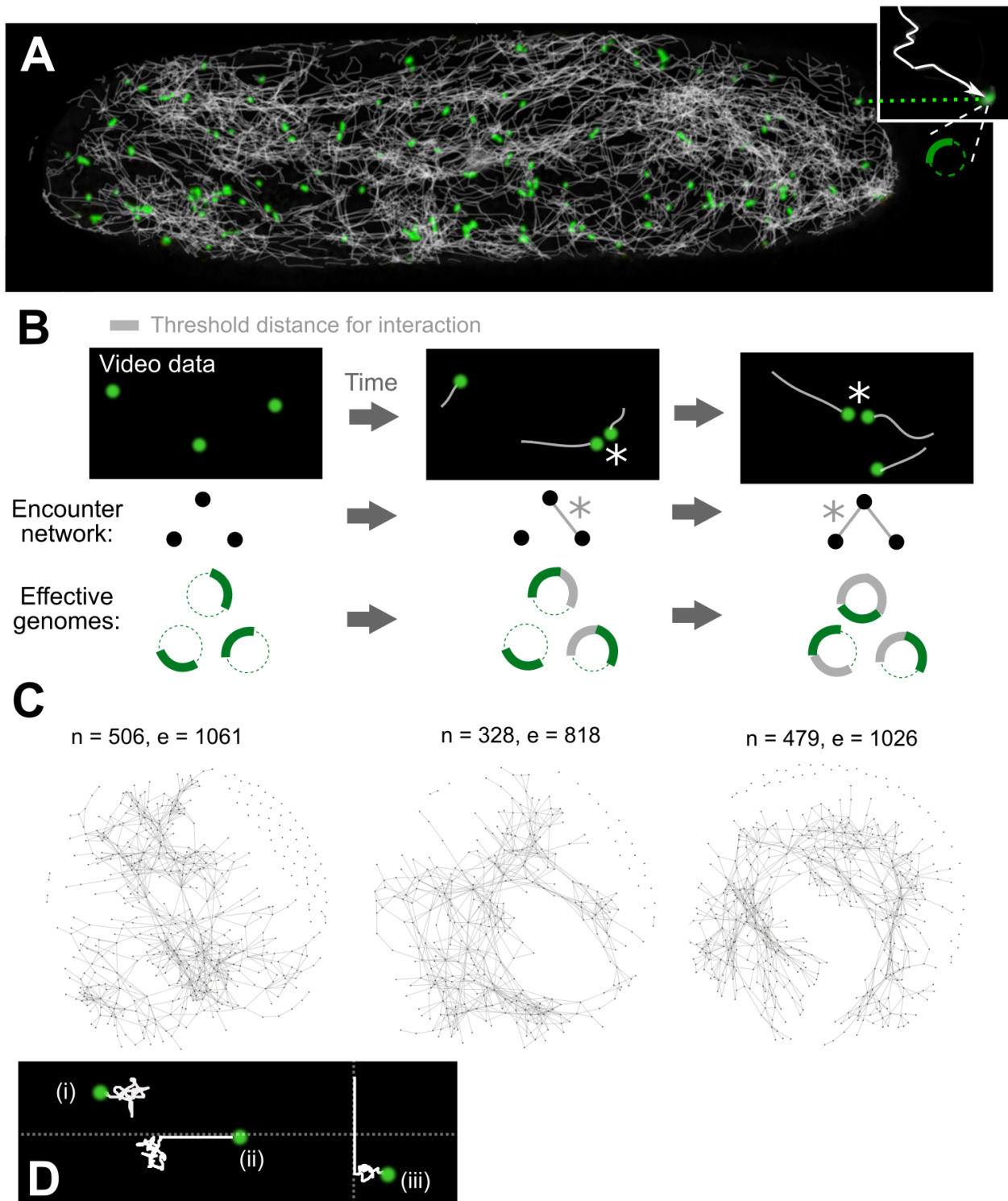
Figure 1: **Characterising mitochondrial encounter networks.** (A) Confocal microscopy with mtGFP *Arabidopsis* [Logan and Leaver, 2000] creates videos of the motion of mitochondria (green) in hypocotyl cells. TrackMate [Tinevez et al., 2017] in Fiji [Schindelin et al., 2012] is used to characterise trajectories (white; individual shown in inset). Individual mitochondria, as illustrated in inset, may only carry a reduced mtDNA molecule encoding a subset (thick line) of the full genome (dashed line). (B) Trajectory sets are interpreted as encounter networks by representing each mitochondrion as a node, and connecting two nodes with an edge if they are ever colocalised within a given threshold distance (∗). Encounters between mitochondria, if they lead to fusion and exchange of mtDNA, can expand the 'effective genome' seen by a mitochondrion, as illustrated: green regions are the mtDNA molecule currently within a mitochondrion, grey regions are those that to which the mitochondrion has been recently exposed. (C) Example encounter networks constructed over a period of 231 seconds n nodes, e edges, cc connected components). (D) Simple physical model used to simulate mitochondrial motion. Model mitochondria may (i) move purely diffusively with constant D; (ii) attach to the cytoskeleton with probability k_{on} and then move ballistically; (iii) detach from the cytoskeleton with probability k_{off} and continue to diffuse.

Our network phrasing is as follows. Allow each node to have two labels G (genome) and H (history), both binary vectors of length L. G describes the set of genetic elements within a node. H describes the set of genetic elements that have been within a node at some point in the past. When a node's H-label contains L elements of value 1, that mitochondrion has been exposed to every genetic element in the full set.

Define an *exchange event* between two nodes a and b, connected by an edge, as follows. The H-label of a acquires a 1 value at every element where the G-label of b is 1. The H-label of b acquires a 1 value at every element where the G-label of a is 1. Then the G-labels of a and b are exchanged. Such an event corresponds to two mitochondria exchanging genetic information, with each being exposed to the genetic information currently in the other.

A given instance of the problem is defined by initial conditions (G-labels for each node) and an adjacency matrix. We are interested in how the H-labels of nodes (the sets of genetic elements that mitochondria have been exposed to) change as the number of exchange events increases. Following the nomenclature in the main text, the bingo score of a node is the proportion of 1s in its H-label $\sum_{i=1}^{L} H_i/L$, and a bingo is scored when this score is 1.
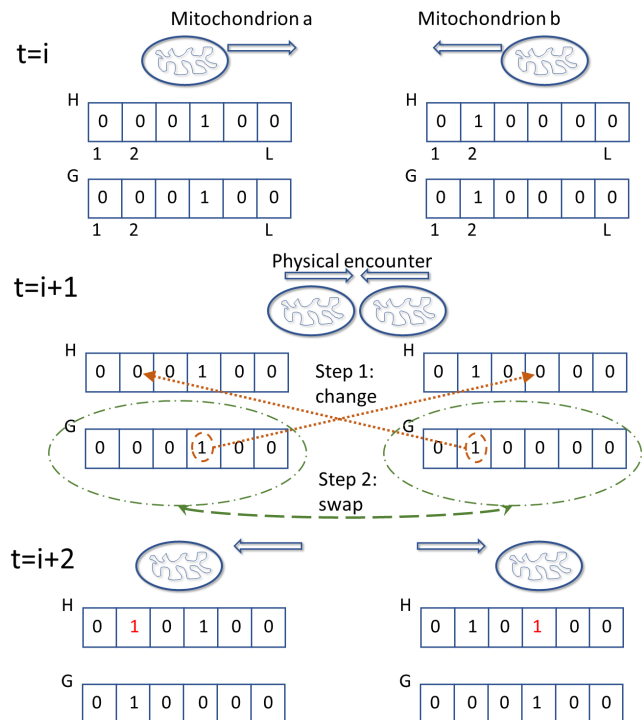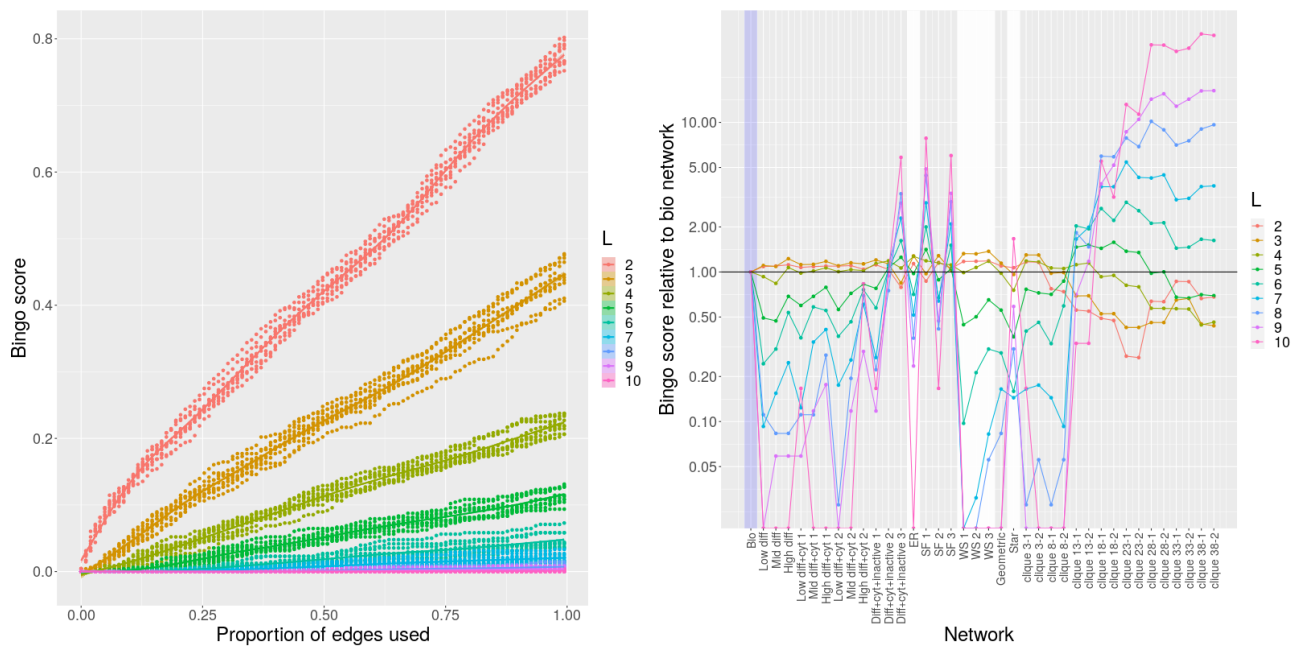

Figure 2: **Effective genome emergence as a network science problem.** Outline of the algorithm modelling genetic exchange on encounter networks. Visual illustration gives an example of the corresponding cell biological process: at time i, two mitochondria with different genomes G move towards each other. At time $i+1$, they undergo a physical encounter and exchange genomes. Their H libraries (set of genes that have been seen at some time point) are updated with the new content their partner has provided. Their G genomes are then exchanged.

– as a network science problem. The biological problem is: how can individual mitochondria become exposed to the full mtDNA genome, given that each may only carry a reduced molecule? We will refer to an 'effective genome' as the set of genes that a mitochondria has been exposed to over time.

We present the specific phrasing of this problem in Fig. 2. Qualitatively, we ask how many genes an individual mitochondrion is exposed to over time, as a function of the proportion of encounters between mitochondria that lead to genetic exchange. We start with some initial state where each mitochondrion contains some or no genetic information, and investigate how information spreads through the population of mitochondria as exchanges between mitochondria occur. This problem shares structural similarities with a wide range of problems in epidemiology [Karp et al., 2000, Moore and Newman, 2000, Kempe et al., 2004, Akdere et al., 2006, Chakrabarti et al., 2008], probability theory (including variants of the coupon collector problem [Flajolet et al., 1992, Newman, 1960]), communication networks and algorithms (including the requirement for every node in the network to acquire required information about the existence of their neighbors [Vasudevan et al., 2009, Ye et al., 2012]), but has some key differences (see Discussion).

For brevity, we refer to this as the *bingo problem*, by analogy with the collection of a set of elements which is built up over time. A node's *bingo score* is the proportion of genetic elements that it has been exposed to over time. A *bingo* occurs when a node has a bingo score of one. This corresponds to a mitochondrion having been exposed to the full set of elements in the genome. An informative summary of a given cell's performance is the proportion p of nodes that have scored bingos (the proportion of mitochondria that have been exposed to a full effective genome).

While a physical encounter does not necessarily imply fusion and exchange of genetic content, it is a

Figure 3: **Potential for genome emergence on *Arabidopsis* encounter networks and on different theoretical network structures.** (A) The 'bingo score' (proportion p of mitochondria that have experienced a full effective genome), as a function of the proportion q of encounter network edges (physical encounters) that allow genetic exchanges. As q increases, genetic information spreads through the mitochondrial population, and more individuals 'collect' the full set of genetic information, increasing p. This increase depends strongly on L, the number of different genetic elements that constitute the full effective genome: higher L means more elements must be 'collected', which requires correspondingly more information exchange. Ten simulations were performed for each L value, using an experimentally-characterised *Arabidopsis* encounter network (see text). (B) Final bingo score p^*, the proportion of mitochondria that have experienced a full effective genome if all encounters allow genetic exchange, is computed for all graph types. This plot shows p^*/p_0^*, this quantity normalised by the value for the biological network structure. For low L, some theoretical networks outperform biology but cliquey networks perform poorly. For high L, the situation is reversed. Traces connecting different network results are drawn to reflect the profile of results for a given L and do not reflect any relationship between different networks. Networks immediately to the right of 'Bio' are encounter networks from physical simulation; others from synthetic construction. Labels: *diff*, diffusion; *cyt*, cytoskeletal motion; *inactive*, stochastic inactivation of mitochondria (modelling entering and leaving the domain); *ER*, Erdős-Rényi; *SF*, scale-free; *WS*, Watts-Strogatz; *clique x-y*, graph with cliques of size x, disconnected if $y = 1$ or connected by a single edge if $y = 2$. Different network classes appear on alternating grey and white backgrounds.

requisite for this exchange. We therefore consider how effective genomes emerge as a changing proportion of encounters are interpreted as leading to exchanges. On one hand, if no encounters lead to exchanges, effective full genomes will never emerge. On the other, if every encounter leads to an exchange, effective full genomes may emerge readily. To characterise this behaviour, we simulated effective genome emergence via the 'bingo' game in Fig. 2. We recorded the proportion p of nodes that have scored a bingo (the proportion of mitochondria that have experienced a full effective genome) as a function of the proportion q of encounters that correspond to an exchange. We increase q following the temporal ordering of encounters in the network.

Intuitively, the dynamics of genome emergence depend strongly on L, the number of different genetic elements that are required to make up a full effective genome (Fig. 3A). For low $L = 2$, effective genomes rapidly emerge with low numbers of interactions, and in the $q = 1$ case where all edges lead to exchange, a majority of mitochondria are able to collect a full effective genome. For higher L, collection becomes increasingly challenging, with only around 10% of mitochondria collecting a full effective genome with $q = 1$ and $L = 5$, and fewer for higher L.

5

Arabidopsis encounter networks support efficient emergence compared to theoretical encounter networks

Having characterised the potential for effective genome emergence on *Arabidopsis* encounter networks, we next asked how these biological networks compared to theoretical alternatives in their capacity to support such emergence. To this end, we investigated the bingo problem on a set of synthetic encounter networks.

For each experimentally-characterised network, we built a range of synthetic networks constrained to have the same numbers of nodes and edges (Fig. S1). Our theoretical networks began with Erdős-Rényi (ER) random topologies ([Erdős and Rényi, 1960]; edges placed between pairs of nodes randomly chosen with uniform probability), scale-free (SF) topologies ([Barabási and Albert, 1999]; edges placed between pairs of nodes randomly chosen with probability proportional to their degree), and Watts-Strogatz (WS) networks ([Watts and Strogatz, 1998, Moore and Newman, 2000]; a 'ring-like' network with subsequent rewiring to reduce networks distances).

We further explored several other network types: geometric random graphs [Penrose, 2003], star graphs, and 'cliquey' graphs. The final class followed our hypothesis that 'cliquiness' in networks would more directly lead to efficient genome emergence, as follows. Cliquey networks consist of cliques (sets of nodes that are all mutually connected) with few or no connections outside each clique. Nodes within cliques can then rapidly assimilate all available genes without risk of 'losing' them to a broader set of partners. We constructed two classes of cliquey network: (i) disconnected cliques of size n and (ii) cliques of size n connected by a single link. In each of these synthetic cases, we specified a number of nodes to match a biologically observed network and padded the network with random edges if necessary to match that network's edge count.

We found that the bingo performance of different networks depends strongly on L, with some networks performing relatively well at $L \leq 3$ (ER, WS, geometric, small cliques) and poorly at $L \geq 4$, and some with the opposite pattern (larger cliques) (Fig. 3B, Fig. S8).

This picture immediately suggests a tension in clique size. Smaller cliques will share information more rapidly. But if a clique is too small, it may not possess all the genes required to accumulate the full set. We found that for $L = 2$, bingo performance was a simple function of clique size, with smaller cliques (down to $n_c = 3$) performing best, and larger cliques (up to $n_c = 38$) performing worst. However, as L increased, this picture became more nuanced. For $L = 3$, the performance of $n_c = 3$ networks was substantially challenged, due to the probability of a clique not possessing a copy of each genetic element. For $L = 3$, larger cliques ($n_c = 8$) performed better, with even larger clique sizes (n_c between 10 and 25) performing best for higher values of $L = 4$ to $L = 6$. Larger cliques $n_c > 30$ performed poorly in most cases, only becoming broadly competitive at high L values.

However, the more striking result was that biological networks and SF networks were the most robust performers. While never being the best performer for a given L, these networks performed much more consistently across a range of different L values (Fig. 3B).

Heterogeneous diffusive and ballistic motion supports efficient effective genome emergence

We next asked which properties of biological mitochondrial motion were responsible for the formation of encounter networks with strong bingo performance. To this end, we considered a simple physical simulation following [Chustecki et al., 2021] (Fig. 1D; Methods). Within the simulation, mitochondria move diffusively, with some probability of attaching to a cytoskeletal strand, whereupon they move ballistically until they detach with some probability. The attachment-detachment probabilities, diffusion constant, and speed when attached to a strand are parameters of the simulation.

Exploring a range of parameters in this model (see Methods), we found that no instance of the diffusive-ballistic model produced encounter networks that could outperform biological networks at bingo. While simulated performance was marginally higher for $L \leq 3$, performance at higher L was substantially lower, only

approaching the biological case for unphysically high values of the diffusion constant and ballistic speed (Fig. 3B). The degree distributions of networks constructed through simulation typically had more limited spread, with fewer nodes of high degree (Fig. S2).

We and others previously observed pronounced inter-mitochondrial heterogeneity in dynamics. Some mitochondria persist in a given cellular region for a long time period, whereas others enter and leave the region, leading to heterogeneity in the time windows for which a given mitochondrion is present. Those individuals present for longer have more opportunity to encounter partners and become highly connected. To model this, we introduced another process in our simulation model, allowing mitochondria to enter and exit the region of observation randomly with given rates (see Methods). As before, we used simulations to produce encounter networks matching the node and edge count of the biological original. We found that these simulated networks, with high diffusion and cytoskeletal motion, more resembled the biological bingo performance (Fig. 3B). Hence, a combination of diffusive and ballistic motion with broader variability in individual behaviour builds a foundation for efficient genome emergence.

To further explore this observation, we next artificially truncated the length of tracked trajectories in the biological data. Unsurprisingly, this led to smaller encounter networks, but also amplified the performance boost of scale-free and beneficially cliquey networks (Fig. S9). This observation supports the picture where a subset of individuals, remaining in the system for a comparatively long time period, accumulate more encounters and thus help facilitate the beneficial exchange of contents.

Network properties linked to efficient effective genome emergence

Given these observations, we next asked whether simple summary statistics of network structure correlated with bingo performance, and hence whether particular structural features might conceivably be selected in cellular control of mitochondrial encounter networks. It may be anticipated that a network's performance at bingo would be related to how rapidly information can be spread through the network. This rapidity is captured by statistics like the global network efficiency $\nu = (n(n-1))^{-1} \sum_{i \neq j \in G} d(i,j)^{-1}$, the sum of the reciprocals of shortest path lengths $d(i,j)$ between all pairs of nodes i and j, normalised by the number of pairs $n(n-1)$. Structural statistics like modularity (which we measure here using the walktrap algorithm [Pons and Latapy, 2006]) and the size and structure of connected components may also be anticipated to play a role (the mean degree, by construction, is equal across all networks compared in an experiment).

However, when exploring bingo behaviour on our synthetic networks, we found that networks with high efficiency, and high values of other intuitively desirable statistics, often do not perform well at bingo (Fig. S4). It is in every node's interest to be the only node connected to as many other sources of information as possible; efficient networks typically connect 'everything to everything'. Other summary statistics also failed to show a tight correlation to bingo performance. While some correlated strongly for a given L (for example, increasing number of connected components decreases performance for $L = 2$), these relationships were typically reversed for different L (increasing number of connected components increases performance for $L = 5$). One suggestive observation is that those networks that perform most consistently – SF and biological networks – have a high degree 'range', here defined as the number of values k for which at least one node in the network has degree k (Fig. S2). This quantity is at least somewhat related to the 'scale-free' nature of a network – degrees spanning a wide range of values – perhaps suggesting the capacity to accumulate information over a diverse ranges of 'scales' of L.

Given this observation, we next considered a more concrete theoretical framework to understand the problem of effective genome emergence – specifically, the coupon collector's problem or CCP [Ferrante and Saltalamacchia, 201 The informal phrasing of the problem is: if each cereal box contains a random coupon, and there are n different types of coupon, how many cereal boxes do I need to buy to collect all n types? The CCP generally describes the process of sampling coupons (which are individual members of a set of coupons $\mathcal{L}$) from a certain number n of 'urns' (entities containing coupons) n. In our system, coupons correspond to individual genome regions (members of the full genome), and urns correspond to mitochondria containing these genes (to further draw
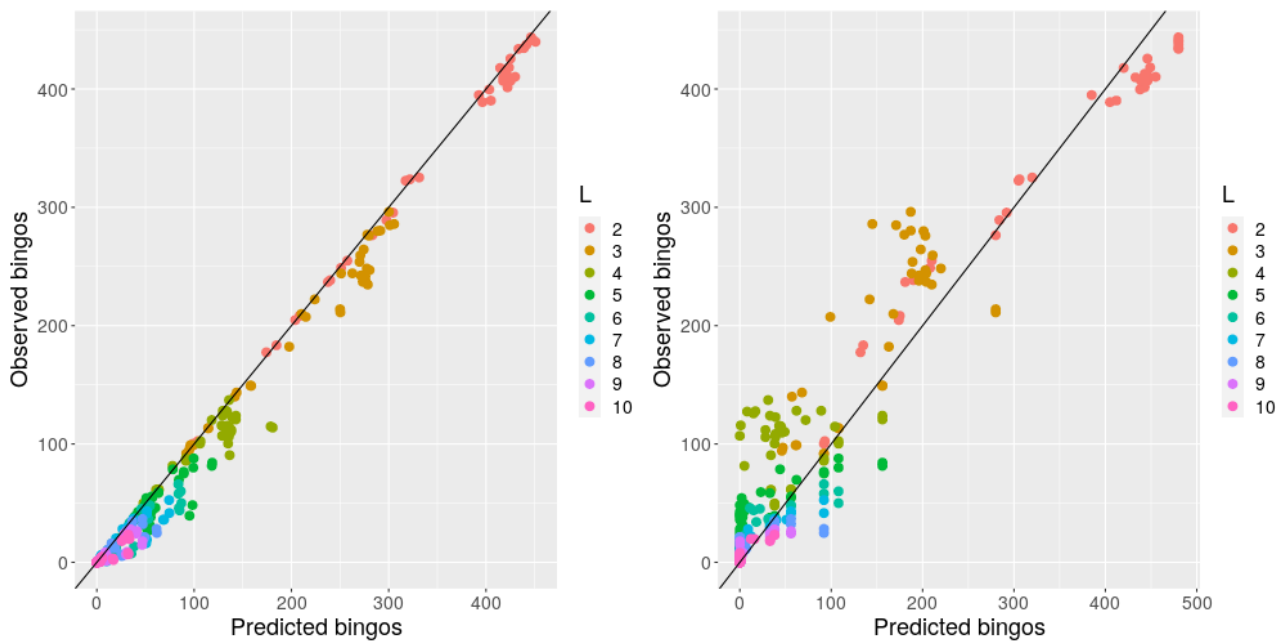
Figure 4: **Analytic results predict bingo performance on biological and synthetic encounter networks.** Predicted (horizontal axis) and observed (vertical axis) number of nodes scoring bingo; each point corresponds to a different network. (left) expected results from Eqn. 1; (right) estimated performance using the threshold value derived from the expected number of encounters needed to fill a bingo. There are 20 repetitions of each bingo game with different L, each value of L is represented with a different colour.

the analogy between the CCP and the bingo game for effective genome, we refer to the visualised glossary in Supp. Fig. S6.) We consider the CCP faced by an individual mitochondrion – a node s in our encounter network. This node begins with its initial gene, and through encounters can draw from each its neighbours (of which there are $deg(s)$). So its total number of draws is $n(s) = deg(s) + 1$, and the number of distinct coupon types to collect is $|\mathcal{L}| = L$.

Study of the CCP has answered many questions about this system – some examples linked to this system appear in Refs. [Flajolet et al., 1992, Adler et al., 2003, Schilling, 2021]. The most central for us is, given n draws, what is the probability of collecting all L coupons? A classical result, outlined in Methods, is that

$$P(s \text{ scores bingo}) = \sum_{j=0}^{L}(-1)^j \binom{L}{j} \left(\frac{L-j}{L}\right)^{n(s)}, \qquad (1)$$

The expected number of neighbours required to score a bingo is also easily derived (see Methods) to be

$$E(n(s)|s \text{ scores bingo}) = LH_L, \qquad (2)$$

where H_L is the Lth harmonic number.

Given this quantity, we are able to characterize and 'predict' the behaviour of a graph structure in bingo, including mitochondrial encounter networks, based on a simple scalar property of the network. Fig. 4 confirms that the bingo game corresponds to the CPP variation described in Eq. 1. We see that the equation predicts the game's outcome for the majority of network topologies and across different values of L; the approximate prediction using the expected value given by Eq. 2 also reasonably predicts the bingo outcome, while requiring only a summary statistic of the whole network.

These insights support the intuitive observation that nodes with degree less than L can never score a bingo, and thus have a purely negative effect on the bingo performance of a network when measured by the proportion of bingo scores. Such nodes, including 'singletons' with degree zero, do occur in our biological encounter networks, because of the limited time window of our observation (see Discussion). To check how much our general results depend on the presence of these low-degree nodes, we artificially removed degree-zero nodes

8

from our biological encounter networks, and re-analysed these 'pruned' networks as above, constructing new synthetic and simulated networks to match the new node and edge counts. We confirmed that networks with 'pruned' and original statistics showed very comparable behaviours, showing that the typically small proportion of singletons does not dramatically influence overall network performance (Fig. S9.)

Presence of 'master circles'

The previous sections considered a population of mitochondria where each mitochondrion begins with one genetic element. An alternative picture in plant biology is that of 'master circles' and 'small circles'. Here, some mitochondria possess full copies of the mtDNA genome, and many mitochondria possess reduced copies. Those with full copies have been pictured as 'genetic vaults' or 'repositories' of genetic information [Logan, 2010, Johnston, 2019].

We next investigated how the presence of master circles influences the emergence of an effective genome across mitochondria. Clearly, the dynamics of the bingo game will differ, because any mitochondrion with a master circle immediately attains a score of one. We explored the behaviour of the system when 1% or 2% of the mitochondrial population was initialised with a master circle, denoting by m the proportion of mtDNA molecules that are master circles. Intuitively, we saw higher bingo scores over time in the cases where more master circles were present (Fig. S7). We also saw a decrease in the scale of differences between networks, with biological and theoretical networks performing more similarly. Interestingly, the presence of master circles induces non-monotonic behaviour of performance with L. In contrast with the $m = 0$ case, where high L values corresponded to maximal cliquey performance and minimal simulation performance, here cliquey performance is maximised and theoretical performance minimised around $L = 5$ or $L = 6$, with the trend reversing at both higher and lower L. The dynamics of bingo is more comparable across different networks for $m > 0$ (Fig. S5), with biological networks performing generally well across all L. Hence, while $m > 0$ makes the bingo problem generally easier for all networks, and biological networks perform correspondingly well.

Discussion

The previous research presenting these encounter networks [Chustecki et al., 2021] hypothesised that the collective dynamics of plant mitochondria allow the cell to balance two priorities. The first is an even physical distribution of mitochondria, ensuring a uniform energy supply, potential for colocalisation with other organelles throughout the cell, and avoiding heterogeneity in concentration of metabolites and signalling molecules. The second 'social' priority is colocalisation of mitochondria to facilitate exchange of genetic information and biomolecules. Here we build on this second priority to show that the topology of encounter networks is capable of facilitating the efficient emergence of an effective genome.

We have shown that the encounter networks observed in plant cells support the emergence of a full 'effective' mtDNA genome in an efficient and distributed way. We showed that biological encounter networks support this process more than a range of synthetic and simulated networks, demonstrating that this process is an analogy of the coupon collector's problem (CCP) and characterising it mathematically.

The above explains some interesting observations from the empirical part of the paper:

Scale-free networks demonstrated good bingo performance across L. This happens because of the well-known heavy tails in their degree distribution. This enables them to perform adequately well for large L and at the same time they are efficient for small L, as well (showing high degree of robustness [Liu et al., 2017]).

Biological networks perform relatively well at bingo. The degree distribution of biological encounter networks is quite similar to the degree distribution of the scale-free networks (for the same given number of nodes and edges). This enables them to have a tail almost the same as the scale-free cases. This is not new, since mitochondria are seen to form *social*-like networking structures [Chustecki et al., 2021] and the results here seem to further validate those findings.

Cliquey networks have an 'unstable' performance for different L. Because for large values of L and big clique size they act as an approximation of a complete graph, which in turn has the maximum degree distribution and hence it is the optimum topology to play bingo (if given of course a much higher number of edges) [Aldous, 1989]. On the other hand, they fail to perform well for small L because they have 'sacrificed' big chunks of the network edges to form the cliques. The exact opposite behaviour is observed for small cliques, where the performance is maximized for low values of L, whereas for $L > 5$ they do poorly.

Other networks (diffusion, random, and so on) do reasonably for small L, but less well for larger ones. Because their degree distribution lacks a tail needed to expect successful bingos. For example, for $L = 8$, 21 neighbors are expected in order to have a complete bingo with high probability.

We have not considered mtDNA replication, degradation, recombination, or other genetic dynamics [Johnston, 2019] in this model. Plant mitochondrial DNA readily recombines (unlike animal mitochondrial DNA), allowing mixing and restructuring of the information shared between mtDNA molecules [Woloszynska, 2009]. Here we only consider the question of mitochondrial access to genetic information, not the population dynamics and/or restructuring of the molecules containing this information. This is a rich topic in itself, addressed by some classical [Atlan and Couvet, 1993, Albert et al., 1996] and some recent [Edwards et al., 2021] theory, and the influence of these physical dynamics of mitochondria on the genetic dynamics of mtDNA is an ongoing topic of research [Mogensen, 1996, Mouli et al., 2009, Poovathingal et al., 2009, Johnston, 2019, Aryaman et al., 2019, Tam et al., 2013, Tam et al., 2015, Hoitzing et al., 2017]. We underline that the details of rates and magnitudes of our proposed mechanism remain hypothetical: although elegant experiments have demonstrated contents exchange and mixing throughout the chondriome [Arimura et al., 2004, Arimura, 2018], the physical and temporal scales of inter-organelle mtDNA exchanges remain, to our knowledge, uncharacterised. Experimental characterisation of these processes will allow parameterisation of our model, which for now demonstrates the range of possible behaviours and general principles without specifying given parameter values.

Like any approach based on imaging, our characterisation of biological encounter networks is subject to some noise. The requirements to image the cell with a fine time resolution (so that mitochondria can be accurately tracked) and with limited laser power (to avoid damaging the cell) limit the resolution of individual frames, and the motion of mitochondria, while largely confined to a 2D plane, can sometimes lead to individuals being lost during the tracking process. This can affect the structure of the subsequent encounter networks. However, the most common issue – a mitochondrion being transiently 'lost' and hence, for example, being represented as two mitochondria (before and after the 'loss') early and late – will generally have the effect of reducing the degree of nodes. This is because the set of encounters of such a mitochondrion will be split between the two individuals. We thus expect the 'true' encounter network to involve more higher-degree nodes, thus supporting the distinction from the synthetic cases with limited degree distributions. On a similar note, our protocol involves imaging over a finite time window. Over time, encounter networks will gain more edges, and it is conceivable that over a long time the networks will come to resemble a complete graph, with every mitochondrion having encountered every other. However, there is another timescale in the system: the timescale on which genetic information is 'forgotten', as protein products expressed from a historically-encountered genome molecule degrade. The system is thus expected to avoid steady state behaviour, and our approach informs about the dynamics that shape the system in a sampled window of this out-of-equilibrium behaviour. Further, plant cells are dynamic systems capable of responding to internal and external stimuli via sensing and feedback control. As such, the topology of a cell's encounter network is not fixed over the lifetime of the cell. Cells may adapt mitochondrial dynamics to favour, for example, 'cliquier' or sparser encounter networks as circumstances demand. The capacity of the cell to control mitochondrial dynamics to optimise mitochondrial exchange, and other priorities, is an exciting target for future work.

Our bingo problem resembles many questions from the field of dynamic networks, found in other fields [Moore and Newman, 2000, Vasudevan et al., 2009, Flajolet et al., 1992, Cao et al., 2018]. One particular feature of our plant system is that information cannot be duplicated (mtDNA molecules are assumed not to replicate over the timescale of these dynamics). Once a mitochondrion has been exposed to an element, it remembers that exposure, but can only pass on the information from that element if it possesses an mtDNA molecule

including it – whereupon it loses that molecule.

In conclusion, we have shown that the dynamic encounter networks of mitochondria in *Arabidopsis* cells have the capacity to support efficient mtDNA complementation, allowing individual mitochondria to 'collect' an effective genome despite only ever carrying a reduced subset. Under several circumstances, this genome emergence seems more efficient in biological networks than in many theoretical cases which may be expected to perform well. This suggests an intriguing hypothesis – that the cellular control of these encounter networks may have evolved to facilitate efficiency genome emergence. If this is indeed the case, plant mitochondrial dynamics represents a 'social network' structure under evolutionary control to fulfil an important cellular function.

Methods

Plant growth. (experimental protocols follow those in Ref. [Chustecki et al., 2021]). Seeds of *Arabidopsis thaliana* with mitochondrial-targeted GFP (kindly provided by Prof. David Logan [Logan and Leaver, 2000]) were surface sterilized in 50% (v/v) household bleach solution for 4 minutes with continual inversion, rinsed three times with sterile water, and plated onto $\frac{1}{2}$ Murashige and Skoog (MS) agar. Plated seeds were stratified in the dark for 2 days at 4°C. Seedlings were grown in 16hr light/8hr dark at 21°C for 4-5 days before use.

Imaging. Prior to mounting, cell walls were stained with 10μM propidium iodide (PI) solution for 3 minutes. Following a protocol modified from [Whelan and Murcha, 2015], full seedlings were mounted in water on microscope slides, with cover slip. Imaging of dynamic systems in living cells is a balance between spatial/temporal resolution and maintaining physiological conditions. To avoid undesirable perturbations to the system including physical and light stress and hypoxia, all imaging was done maintaining low laser intensities and within at most 10 minutes of mounting to minimise the effects of physical stress and hypoxia (Prof Markus Schwarzländer, personal communication).

A Zeiss 710 laser scanning confocal microscope was used to capture time lapse images. To test robustness of the imaging protocol, a Zeiss 900 with AiryScan 2 detector was also used for several identically prepared samples, with no differences between summary statistics collected from these samples and those from the 710 beyond natural variability. For cellular characterisation we used excitation wavelength 543nm, detection range 578-718nm for both chlorophyll autofluorescence (peak emission 679.5nm) and for PI (peak emission 648nm). For mitochondrial capture we used excitation wavelength 488nm, detection range 494-578nm for GFP (peak emission 535.5nm). Videos were 231 seconds long, with a frame interval of 1.94 seconds, and a resolution (after scaling for standardisation) of 0.2 μm per pixel.

Video analysis. Individual cells were cropped from the acquired video data using the cell wall PI signal using Fiji (ImageJ) [Schindelin et al., 2012]. The size of each video was scaled to the universal length scale 5.0 pixels/μm. We then extracted individual mitochondrial trajectories from the acquired video data using TrackMate [Tinevez et al., 2017]. Typical settings used were application of the LoG Detector filter with a blob diameter of 1μm and threshold of 2-7, filters were set on spot quality if deemed necessary. The Simple LAP Tracker was run with a linking max distance of 4μm, gap-closing distance of 5μm and gap-closing max frame gap of 2 frames. In each case we visually confirmed that individual mitochondria were appropriately highlighted and that tracks were well captured, editing occasional tracks where necessary. XML output from TrackMate was converted to adjacency matrices using custom code (see below).

Null model networks. We constructed several theoretical models for network structure, each with *n* nodes and *e* edges. First, Erdős-Rényi (ER) random networks [Erdős and Rényi, 1960] were constructed by randomly choosing two non-identical nodes *a* and *b*, each with probability $1/n$, and creating an edge between them, repeating until *e* edges were created.

Second, scale-free (SF) networks [Barabási and Albert, 1999] were constructed by randomly choosing nodes with probability $1/deg(a_i) + 1/\sum_j 1/deg(a_j) + 1$. This procedure was repeated *e* times, with degree updated each time, for the basic network (i). Variations of scale-free networks were created in two ways. For (ii), beginning with a linear network where an edge connects each a_i and a_{i+1}, then proceeding as in (i), thus

11

enforcing connectivity. For (iii), a preferential attachment process was performed for each of the n nodes, where a node is connected to a partner a with probability $1/deg(a_i) + 1/\sum_j 1/deg(a_j) + 1$, where the sum j is over nodes added so far to the network. Extra edges are then added as in (i).

Third, Watts-Strogatz (WS) networks [Watts and Strogatz, 1998] were constructed as follows. Compute the mean degree $k = n/e$. Label each of n nodes with successive integers. For each node i, draw $k_i = \lceil k \rceil$ or $\lfloor k \rfloor$ randomly with relative probabilities $\lceil k \rceil - k$ and $k - \lfloor k \rfloor$. If k_i is even, connect i to the $k_i/2$ nodes immediately before it and the $k_i/2$ nodes immediately after it in sequence. If k_i is odd, connect to $(k_i + 1)/2$ 'before' nodes and $(k_i - 1)/2$ 'after' nodes with probability $\frac{1}{2}$, or vice versa with probability $\frac{1}{2}$. For all edges linking i to a node with label $> i$, change the target node with probability β to a different node $\neq i$.

Fourth, 'cliquey' networks were constructed. Given a clique size c and constraints on n and e, the number of cliques allowed was computed as $n_c = \min(\lfloor n/c \rfloor, \lfloor e/(c(c+1)/2) \rfloor)$. The n nodes were partitioned into n_c cliques with edges between each pair of nodes within each clique. These cliques were then either (i) left disconnected; (ii) connected with a single edge linking two cliques; (iii) left disconnected but padded with randomly placed edges to reach e total; (iv) connected with a single edge linking two cliques then padded.

Fifth, geometric random graphs (GRGs) were constructed by placing n points – each representing a node – in the unit square, and progressively adding edges between the two disconnected nodes with the shortest distance between their corresponding points, until e edges existed. Finally, the star graph with n nodes was constructed by connecting $n - 1$ nodes to a central node, then adding random edges until e edges existed.

Model networks based on physical simulation. Synthetic encounter networks were constructed based on physical simulation of model mitochondrial dynamics using custom code in C (see below). As we are free to set length and time units in our simulation, we use $1\mu m$ as the unit of length and set one discrete simulation timestep equivalent to $1s$. n agents were simulated in a model cell, a 2D rectangular domain with reflecting boundary conditions at $x = 0$, $x = 100\mu m$, $y = 0$, $y = 30\mu m$, to model the geometry observed in our experimental observations of hypocotyl cells [Chustecki et al., 2021]. Cytoskeleton strands are modelled as crossing the cell at constant x (horizontal) and at constant y (vertical). Each agent could, at any time point, be detached or attached to the cytoskeleton. If detached, each timestep, agents were moved according to a normal kernel with standard deviation $2D\mu m^2 s^{-1}$, so that $D\mu m^2 s^{-1}$ is the diffusion constant. When first attached, an agent is assigned a velocity vector: while attached, that agent moves by that vector each timestep. The velocity vector is randomly chosen on attachment and may be in the $+x$, $-x$, $+y$, or $-y$ direction, and has magnitude $V\mu m s^{-1}$. Each timestep, detached agents become attached with probability k_{on}, and attached agents become detached with probability k_{off}, corresponding to rates of $k_{on/off}s^{-1}$. When two agents were present within a distance $1.6\mu m$ of each other, an edge corresponding to the pair was added to the encounter network (if not already present). The physical simulation proceeded until e edges were present.

Characteristic values observed experimentally are $D \simeq 0.1\mu m^2 s^{-1}$ and $V \simeq 1\mu m s^{-1}$ [Chustecki et al., 2021]. In our simulations we explored one order of magnitude either side of these values, using $D = (0.02, 0.1, 1)\mu m^2 s^{-1}$ and $V = (0.1, 1, 10)\mu m s^{-1}$. We explored (k_{on}, k_{off}) pairs of $(0, 0)s^{-1}$ (no cytoskeletal motion), $(0.1, 0.1)s^{-1}$, and $(0.5, 0.1)s^{-1}$.

Entry and exit of individual organelles into the system was modelled by switching individuals between 'active' and 'inactive' states. Active mitochondria behave as above and interact; inactive mitochondria remain static and do not contribute to any encounters, remaining effectively invisible (thus having exited the system). When this feature was used in simulations, activation and inactivation of individuals were stochastic events with rates $\rho_{on} = 0.01s^{-1}$ and $\rho_{off} = 0.1s^{-1}$ respectively, leading to a mean of 10% active mitochondria at a given time.

Coupon collector's problem. Consider the different patterns of coupons that can be acquired through n draws. There are L^n possible patterns, which we assume all arise with equal probability. We require the probability of obtaining a pattern in which each of the L coupons is present. To get this we use the inclusion-exclusion principle.

We first write down the probability of obtaining a pattern that is compatible with there being an 'alphabet' of L coupons. The probability of a single draw being compatible with an alphabet of L coupons is $L/L = 1$, so we

begin with a probability of $1^n = 1$. We need to deduct the probability of obtaining a pattern that is compatible with there being an alphabet of $L - 1$ coupons, because every such pattern cannot feature all L coupons. The probability of an individual draw not obtaining a given coupon l is $(L - 1)/L$, so considering each $l \in \mathcal{L}$ we obtain $L \times ((L - 1)/L)^n$.

However, we have now over-counted patterns that are compatible with an even smaller alphabet size of $L - 2$. So we need to add back the patterns that we have missed. The probability of an individual draw not obtaining either of a given pair of coupons (l_1, l_2) is $(L - 2)/L$, so considering each pair of coupons $(l_1, l_2) \in \mathcal{L}$ we have $\binom{L}{2} \times ((L - 2)/L)^n$.

However, we have now over-counted patterns that are compatible with an even smaller alphabet size of $L - 3$. We thus need to consider triplets of excluded coupons, and so on. The process continues iteratively, alternating between adding and subtracting terms (including and excluding) until we reach L terms. From the above it should be clear that the final form is

$$P(\text{bingo}) \quad = \quad 1 - L\left(\frac{L-1}{L}\right)^n + \binom{L}{2}\left(\frac{L-2}{L}\right)^n - ... \tag{3}$$

$$= \quad \sum_{j=0}^{L}(-1)^j \binom{L}{j}\left(\frac{L-j}{L}\right)^n \tag{4}$$

For example, consider $n = L = 3$. Write the 24 three-character strings of length 3 for the set of patterns: AAA, AAB, ... At the first step we include them all. The next step counts all the strings that do not contain A, all those that do not contain B, and all those that do not contain C. Hence, we remove BBC, BCC, and so on – but AAA, BBB, and CCC each get double-counted (once for each coupon they do not contain). The third step recounts all the strings that do not contain A or B, those that do not contain B or C, and those that do not contain A or C, which are exactly those three strings we previously double-counted. As $L = 3$, this is our final step, and we have successfully retained only those strings in which all coupons feature.

The expected number of draws required for a bingo is easier to compute. If we have collected c coupons, the probability of the next draw obtaining an unseen coupon is $(L - c)/L$. Assuming that draws are Bernoulli trials, a geometric distribution describes the behaviour of the system, giving a mean number of draws $L/(L - c)$ required for the next unseen coupon. The expected overall number is then $\sum_{c=0}^{L} L/(L - c) = L(1/L + 1/(L - 1) + ... + 1/1) = LH_L$.

It is important to note that the above equations assume an equal probability for each draw, as this is the case in the bingo game where L values are initially scattered uniformly across the mitochondria/nodes. The influence of the coupon distribution for the same problem is beyond the scope of this work and is still an open topic of active research. The interested reader is referred to the work of Shilling [Schilling, 2021].

Source code and data availability. All source code is available at `github.com/StochasticBiology/mito-network-sharing`. We used TrackMate [Tinevez et al., 2017] in Fiji [Schindelin et al., 2012] for video analysis, C with the igraph library [Csardi and Nepusz, 2006] for network generation and simulation, and R [R Core Team, 2017] with libraries igraph [Csardi and Nepusz, 2006], brainGraph [Watson, 2020], XML [Temple Lang, 2020], ggplot2 [Wickham, 2016], GGally [Schloerke et al., 2021], and gridExtra [Auguie, 2017] for data curation and visualisation.

Acknowledgments

Schwarzländer for advice with design and analysis of imaging experiments.

References

[Adler et al., 2003] Adler, I., Oren, S., and Ross, S. M. (2003). The coupon-collector's problem revisited. *Journal of Applied Probability*, 40(2):513–518.

[Akdere et al., 2006] Akdere, M., Bilgin, C. C., Gerdaneri, O., Korpeoglu, I., Ulusoy, Ö., and Cetintemel, U. (2006). A comparison of epidemic algorithms in wireless sensor networks. *Computer Communications*, 29(13-14):2450–2457.

[Albert et al., 1996] Albert, B., Godelle, B., Atlan, A., De Paepe, R., and Gouyon, P. H. (1996). Dynamics of plant mitochondrial genome: model of a three-level selection process. *Genetics*, 144(1):369–382.

[Aldous, 1989] Aldous, D. (1989). An introduction to covering problems for random walks on graphs. *Journal of Theoretical Probability*, 2(1):87–89.

[Arimura, 2018] Arimura, S.-i. (2018). Fission and fusion of plant mitochondria, and genome maintenance. *Plant physiology*, 176(1):152–161.

[Arimura et al., 2004] Arimura, S.-i., Yamamoto, J., Aida, G. P., Nakazono, M., and Tsutsumi, N. (2004). Frequent fusion and fission of plant mitochondria with unequal nucleoid distribution. *Proceedings of the National Academy of Sciences*, 101(20):7805–7808.

[Aryaman et al., 2019] Aryaman, J., Bowles, C., Jones, N. S., and Johnston, I. G. (2019). Mitochondrial network state scales mtdna genetic dynamics. *Genetics*, 212(4):1429–1443.

[Atlan and Couvet, 1993] Atlan, A. and Couvet, D. (1993). A model simulating the dynamics of plant mitochondrial genomes. *Genetics*, 135(1):213–222.

[Auguie, 2017] Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. R package version 2.3.

[Barabási and Albert, 1999] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *science*, 286(5439):509–512.

[Cao et al., 2018] Cao, Z., Gu, Z., Wang, Y., and Cui, H. (2018). Panacea: A low-latency, energy-efficient neighbor discovery protocol for wireless sensor networks. In *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE.

[Chakrabarti et al., 2008] Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J., and Faloutsos, C. (2008). Epidemic thresholds in real networks. *ACM Transactions on Information and System Security (TISSEC)*, 10(4):1–26.

[Chustecki et al., 2021] Chustecki, J. M., Gibbs, D. J., Bassel, G. W., and Johnston, I. G. (2021). Network analysis of arabidopsis mitochondrial dynamics reveals a resolved tradeoff between physical distribution and social connectivity. *Cell systems*.

[Cohen et al., 2018] Cohen, S., Valm, A. M., and Lippincott-Schwartz, J. (2018). Interacting organelles. *Current opinion in cell biology*, 53:84–91.

[Csardi and Nepusz, 2006] Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, Complex Systems:1695.

[Edwards et al., 2021] Edwards, D. M., Røyrvik, E. C., Chustecki, J. M., Giannakis, K., Glastad, R. C., Radzvilavicius, A. L., and Johnston, I. G. (2021). Avoiding organelle mutational meltdown across eukaryotes with or without a germline bottleneck. *PLoS biology*, 19(4):e3001153.

[El Zawily et al., 2014] El Zawily, A. M., Schwarzländer, M., Finkemeier, I., Johnston, I. G., Benamar, A., Cao, Y., Gissot, C., Meyer, A. J., Wilson, K., Datla, R., et al. (2014). FRIENDLY regulates mitochondrial distribution, fusion, and quality control in Arabidopsis. *Plant physiology*, 166(2):808–828.

[Erdős and Rényi, 1960] Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60.

[Ferrante and Saltalamacchia, 2014] Ferrante, M. and Saltalamacchia, M. (2014). The coupon collector's problem. *Materials matemàtics*, pages 0001–35.

[Flajolet et al., 1992] Flajolet, P., Gardy, D., and Thimonier, L. (1992). Birthday paradox, coupon collectors, caching algorithms and self-organizing search. *Discrete Applied Mathematics*, 39(3):207–229.

[Hoitzing et al., 2017] Hoitzing, H., Johnston, I. G., and Jones, N. S. (2017). Stochastic Models for Evolving Cellular Populations of Mitochondria: Disease, Development, and Ageing. In *Stochastic Processes, Multiscale Modeling, and Numerical Methods for Computational Cellular Biology*, pages 287–314. Springer.

[Johnston, 2019] Johnston, I. G. (2019). Tension and resolution: dynamic, evolving populations of organelle genomes within plant cells. *Molecular plant*, 12(6):764–783.

[Karp et al., 2000] Karp, R., Schindelhauer, C., Shenker, S., and Vocking, B. (2000). Randomized rumor spreading. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 565–574. IEEE.

[Kempe et al., 2004] Kempe, D., Kleinberg, J., and Demers, A. (2004). Spatial gossip and resource location protocols. *Journal of the ACM (JACM)*, 51(6):943–967.

[Liu et al., 2017] Liu, J., Zhou, M., Wang, S., and Liu, P. (2017). A comparative study of network robustness measures. *Frontiers of Computer Science*, 11(4):568–584.

[Liu et al., 2009] Liu, X., Weaver, D., Shirihai, O., and Hajnóczky, G. (2009). Mitochondrial kiss-and-run: interplay between mitochondrial motility and fusion–fission dynamics. *The EMBO journal*, 28(20):3074–3089.

[Logan, 2006] Logan, D. C. (2006). The mitochondrial compartment. *Journal of experimental botany*, 57(6):1225–1243.

[Logan, 2010] Logan, D. C. (2010). Mitochondrial fusion, division and positioning in plants.

[Logan and Leaver, 2000] Logan, D. C. and Leaver, C. J. (2000). Mitochondria-targeted gfp highlights the heterogeneity of mitochondrial shape, size and movement within living plant cells. *Journal of Experimental Botany*, 51(346):865–871.

[Mogensen, 1996] Mogensen, H. L. (1996). The hows and whys of cytoplasmic inheritance in seed plants. *American Journal of Botany*, 83(3):383–404.

[Moore and Newman, 2000] Moore, C. and Newman, M. E. (2000). Epidemics and percolation in small-world networks. *Physical Review E*, 61(5):5678.

[Mouli et al., 2009] Mouli, P. K., Twig, G., and Shirihai, O. S. (2009). Frequency and selectivity of mitochondrial fusion are key to its quality maintenance function. *Biophysical journal*, 96(9):3509–3518.

[Newman, 1960] Newman, D. J. (1960). The double dixie cup problem. *The American Mathematical Monthly*, 67(1):58–61.

[Penrose, 2003] Penrose, M. (2003). *Random Geometric Graphs*. Oxford Studies in Probability. Oxford University Press.

[Picard and Sandi, 2021] Picard, M. and Sandi, C. (2021). The social nature of mitochondria: Implications for human health. *Neuroscience & Biobehavioral Reviews*, 120:595–610.

[Pons and Latapy, 2006] Pons, P. and Latapy, M. (2006). Computing communities in large networks using random walks. In *J. Graph Algorithms Appl*. Citeseer.

[Poovathingal et al., 2009] Poovathingal, S. K., Gruber, J., Halliwell, B., and Gunawan, R. (2009). Stochastic drift in mitochondrial dna point mutations: a novel perspective ex silico. *PLoS computational biology*, 5(11):e1000572.

[Preuten et al., 2010] Preuten, T., Cincu, E., Fuchs, J., Zoschke, R., Liere, K., and Börner, T. (2010). Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. *The Plant Journal*, 64(6):948–959.

[R Core Team, 2017] R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[Schilling, 2021] Schilling, J. (2021). Results and conjectures on the role of the uniform distribution in the coupon collector's problem with group drawings. *Information Processing Letters*, 169:106112.

[Schindelin et al., 2012] Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7):676–682.

[Schloerke et al., 2021] Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., Elberg, A., and Crowley, J. (2021). *GGally: Extension to 'ggplot2'*. R package version 2.1.2.

[Seguí-Simarro and Staehelin, 2009] Seguí-Simarro, J. M. and Staehelin, L. A. (2009). Mitochondrial reticulation in shoot apical meristem cells of Arabidopsis provides a mechanism for homogenization of mtDNA prior to gamete formation. *Plant signaling & behavior*, 4(3):168–171.

[Takanashi et al., 2006] Takanashi, H., Arimura, S.-i., Sakamoto, W., and Tsutsumi, N. (2006). Different amounts of dna in each mitochondrion in rice root. *Genes & genetic systems*, 81(3):215–218.

[Tam et al., 2013] Tam, Z. Y., Gruber, J., Halliwell, B., and Gunawan, R. (2013). Mathematical modeling of the role of mitochondrial fusion and fission in mitochondrial dna maintenance. *PloS one*, 8(10):e76230.

[Tam et al., 2015] Tam, Z. Y., Gruber, J., Halliwell, B., and Gunawan, R. (2015). Context-dependent role of mitochondrial fusion-fission in clonal expansion of mtdna mutations. *PLoS computational biology*, 11(5):e1004183.

[Temple Lang, 2020] Temple Lang, D. (2020). *XML: Tools for Parsing and Generating XML Within R and S-Plus*. R package version 3.99-0.3.

[Tinevez et al., 2017] Tinevez, J.-Y., Perry, N., Schindelin, J., Hoopes, G. M., Reynolds, G. D., Laplantine, E., Bednarek, S. Y., Shorte, S. L., and Eliceiri, K. W. (2017). Trackmate: An open and extensible platform for single-particle tracking. *Methods*, 115:80–90.

[Valm et al., 2017] Valm, A. M., Cohen, S., Legant, W. R., Melunis, J., Hershberg, U., Wait, E., Cohen, A. R., Davidson, M. W., Betzig, E., and Lippincott-Schwartz, J. (2017). Applying systems-level spectral imaging and analysis to reveal the organelle interactome. *Nature*, 546(7656):162–167.

[Vasudevan et al., 2009] Vasudevan, S., Towsley, D., Goeckel, D., and Khalili, R. (2009). Neighbor discovery in wireless networks and the coupon collector's problem. In *Proceedings of the 15th annual international conference on Mobile computing and networking*, pages 181–192.

[Watson, 2020] Watson, C. G. (2020). *brainGraph: Graph Theory Analysis of Brain MRI Data*. R package version 3.0.0.

[Watts and Strogatz, 1998] Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of small-worldnetworks. *nature*, 393(6684):440–442.

[Whelan and Murcha, 2015] Whelan, J. and Murcha, M. W. (2015). *Plant mitochondria: methods and protocols*. Springer.

[Wickham, 2016] Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

[Williams and George, 2019] Williams, C. F. and George, C. H. (2019). Connect and conquer: collectivized behavior of mitochondria and bacteria. *Frontiers in physiology*, 10:340.

[Woloszynska, 2009] Woloszynska, M. (2009). Heteroplasmy and stoichiometric complexity of plant mitochondrial genomesthough this be madness, yet there's method in't. *Journal of experimental botany*, 61(3):657–671.

[Ye et al., 2012] Ye, F., Roy, S., and Wang, H. (2012). Efficient data dissemination in vehicular ad hoc networks. *IEEE Journal on Selected Areas in Communications*, 30(4):769–779.
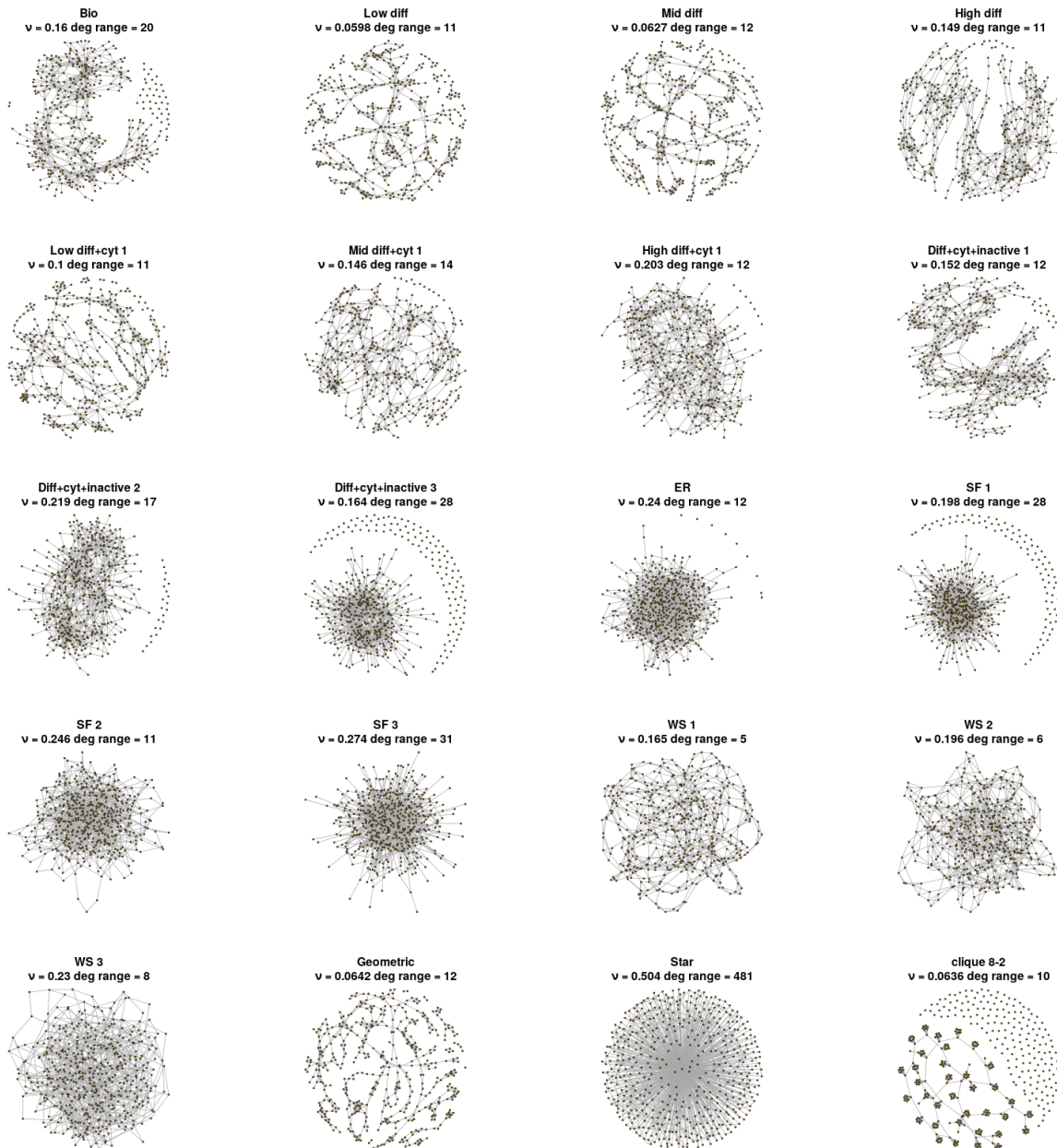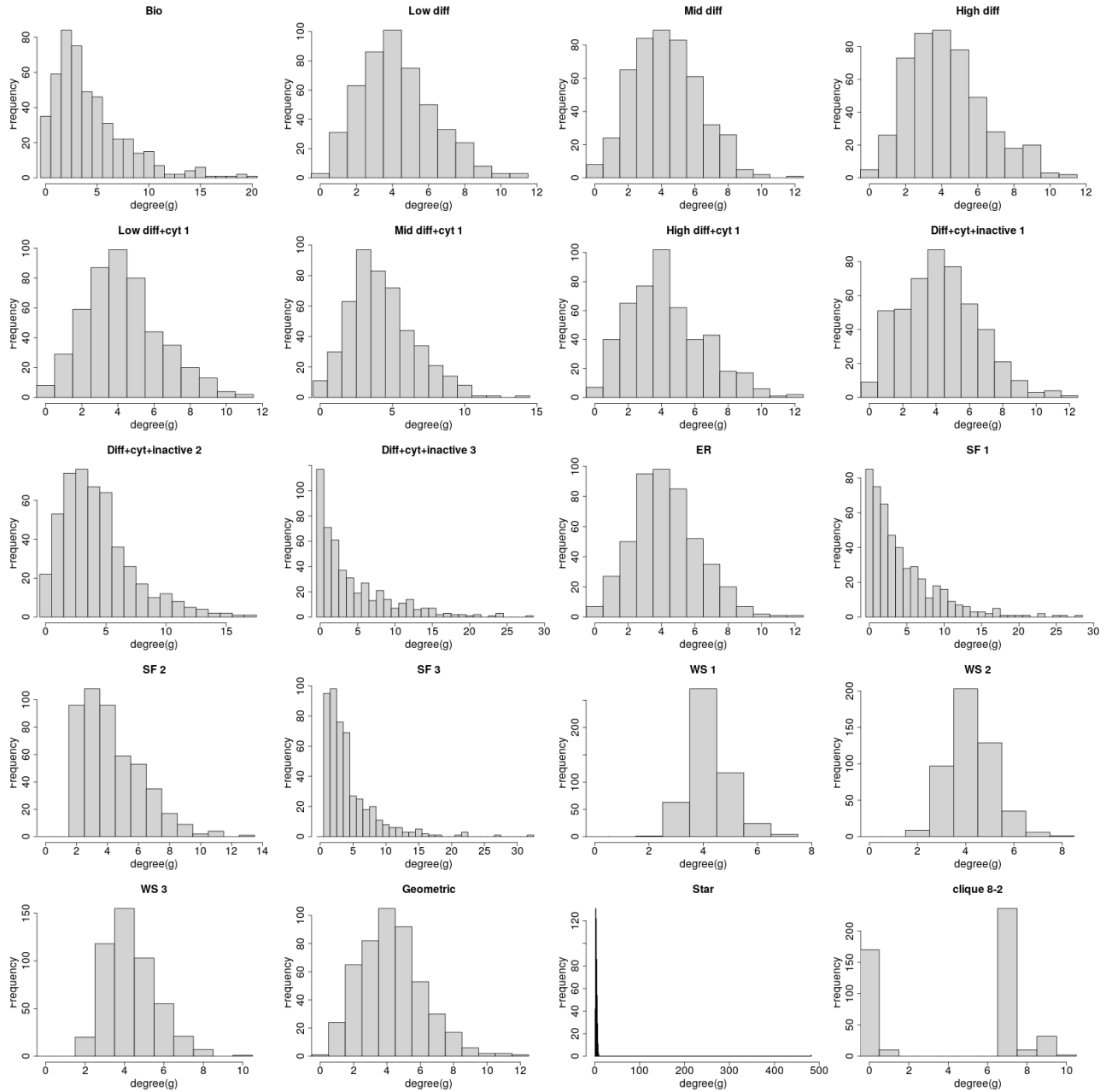
Supplementary Information

Figure S1: **Comparison of encounter networks from experiment, simulation of mitochondrial dynamics, and general theory.** Visualisations of network structures from the different construction protocols described in Methods, matching (as closely as possible) the statistics of the biological ('Bio') network. One representative 'cliquey' network structure is shown; abbreviations are ER (Erdős-Rényi), SF (scale-free), WS (Watts-Strogatz). Network statistics are ν, global efficiency; and *deg range*, range of degree distribution.

Figure S2: **Degree distributions of encounter networks from experiment, simulation of mitochondrial dynamics, and general theory.** Degree distributions for the ensemble of graphs in Fig. S1. One representative 'cliquey' network structure is shown; abbreviations are ER (Erdős-Rényi), SF (scale-free), WS (Watts-Strogatz).
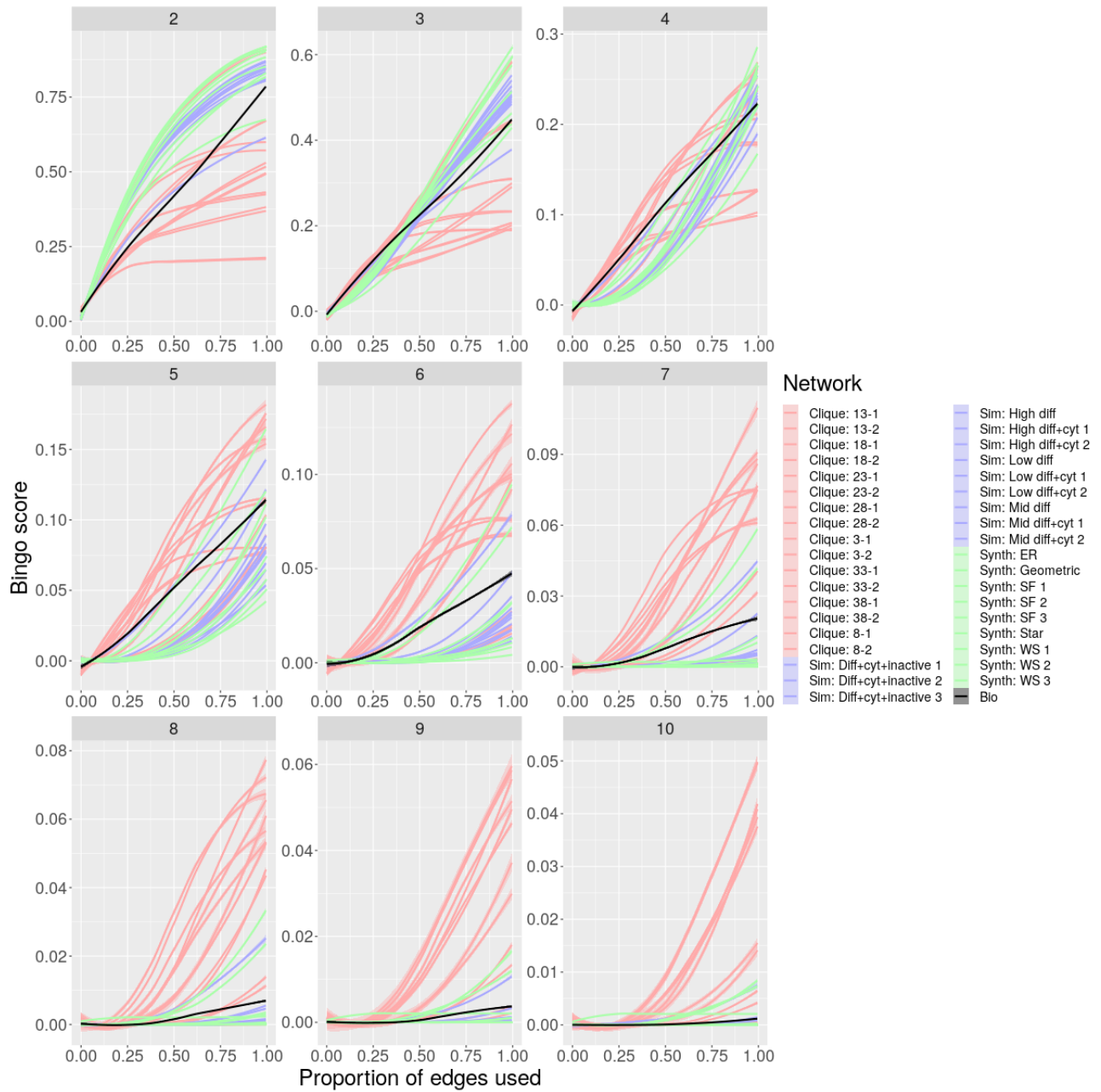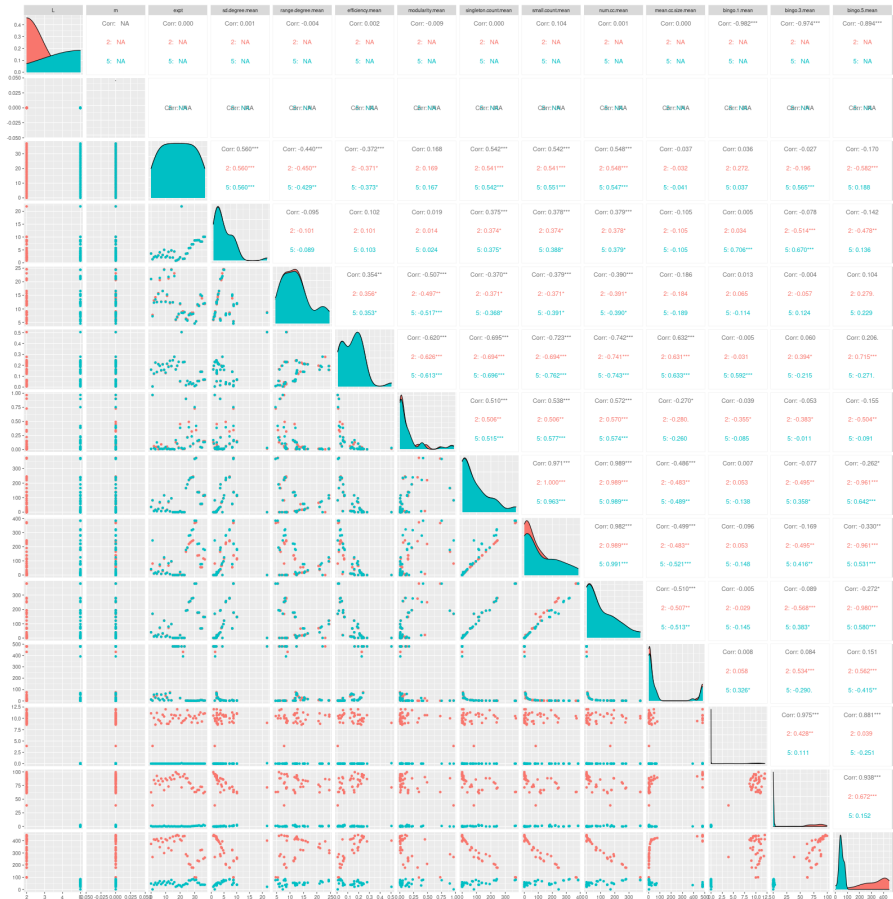
Figure S3: **Bingo dynamics on different networks for different** L **with** $m = 0$. Behaviour of bingo score p with proportion of edges q used for genetic exchange, arranged for a range of synthetic networks and their biological partner, for different L. Traces are coloured by the general class of synthetic network. Traces are LOESS fits to $n = 10$ simulations for each case.

Figure S4: **Network statistics and bingo performance.** Correlations between network statistics and bingo performance, for $L = 2$ (red) and $L = 5$ (blue), with scatter plots under the diagonal, Pearson coefficients above the diagonal (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$), and histograms of the statistic on the diagonal. Each point is a mean value for a different class of network, taken over 10 generated instances. Labels: *sd.degree* and *range.degree*, degree distribution standard deviation and range; *efficiency*, global network efficiency; *modularity*, network modularity measured using the walktrap algorithm [Pons and Latapy, 2006]; *singleton.count*, number of degree-zero nodes; *small.count*, number of components with size $< L$; *num.cc* and *mean.cc.size*, number and mean size of connected components; *bingo.1/3/5*, bingo score when proportion 0.01/0.1/1 of edges are used for genetic exchange. Although some statistics correlate with bingo performance for a given L, little correlation across L values is visible.
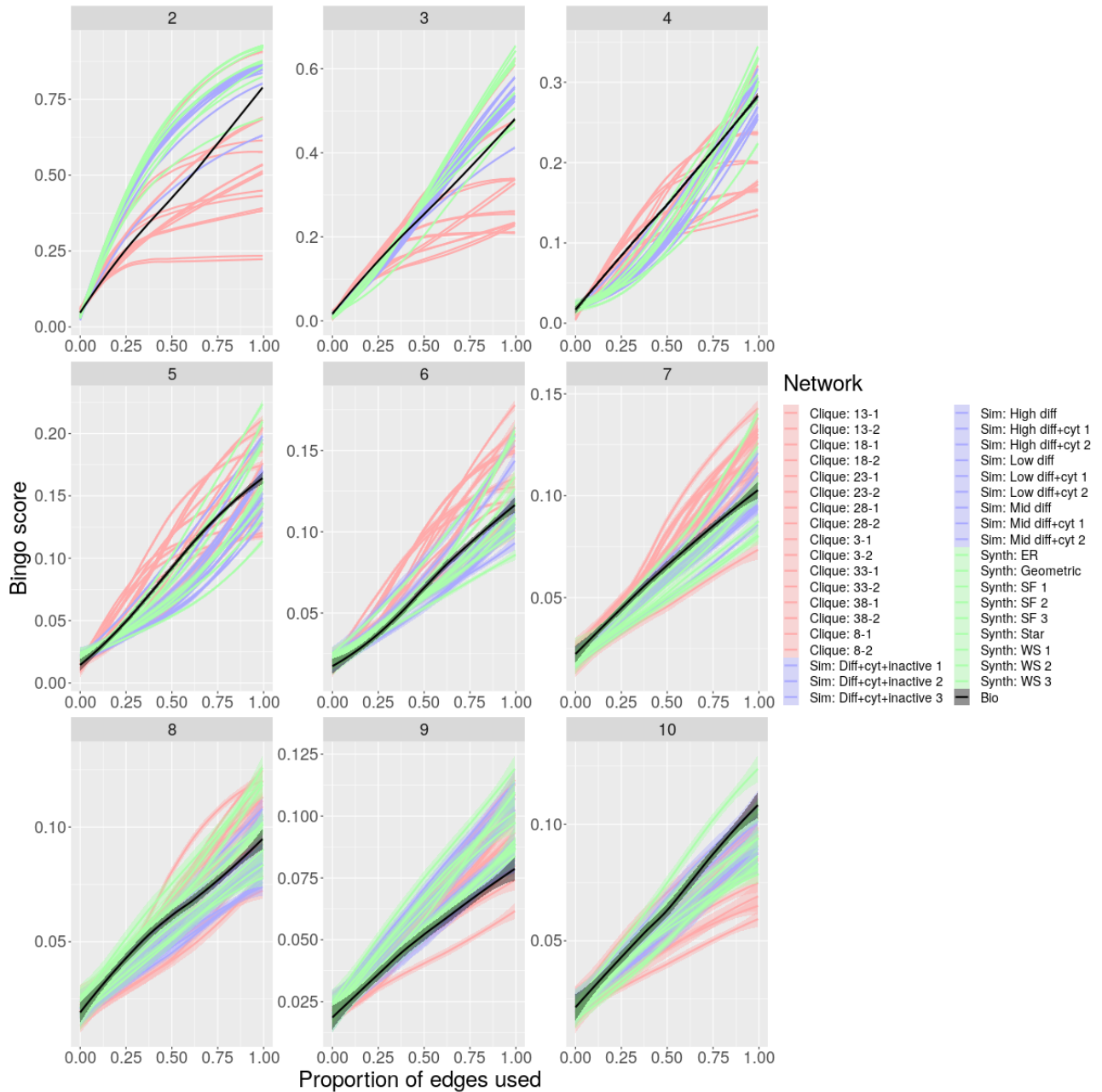
Figure S5: **Bingo dynamics on different networks for different** L **with** $m = 0.02$**.** Analogous to Fig. S3, behaviour of bingo score p with proportion of edges q used for genetic exchange, arranged for a range of synthetic networks and their biological partner, for different L and a master circle proportion of $m = 0.02$. Traces are coloured by the general class of synthetic network. Traces are LOESS fits to $n = 10$ simulations for each case.
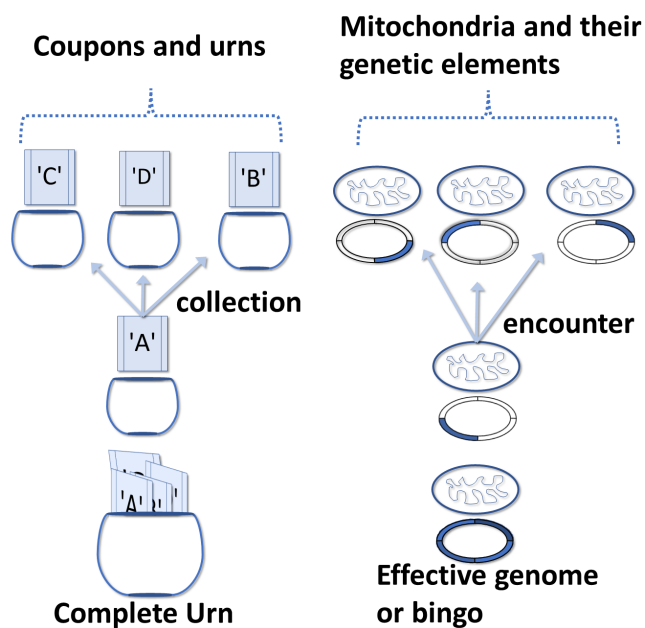
Figure S6: **The analogy between the CCP and bingo.** It is shown and explained the shared terminology between the two concepts and how the coupon collection corresponds to the the assembly of effective genome through encounters with partial genome elements.
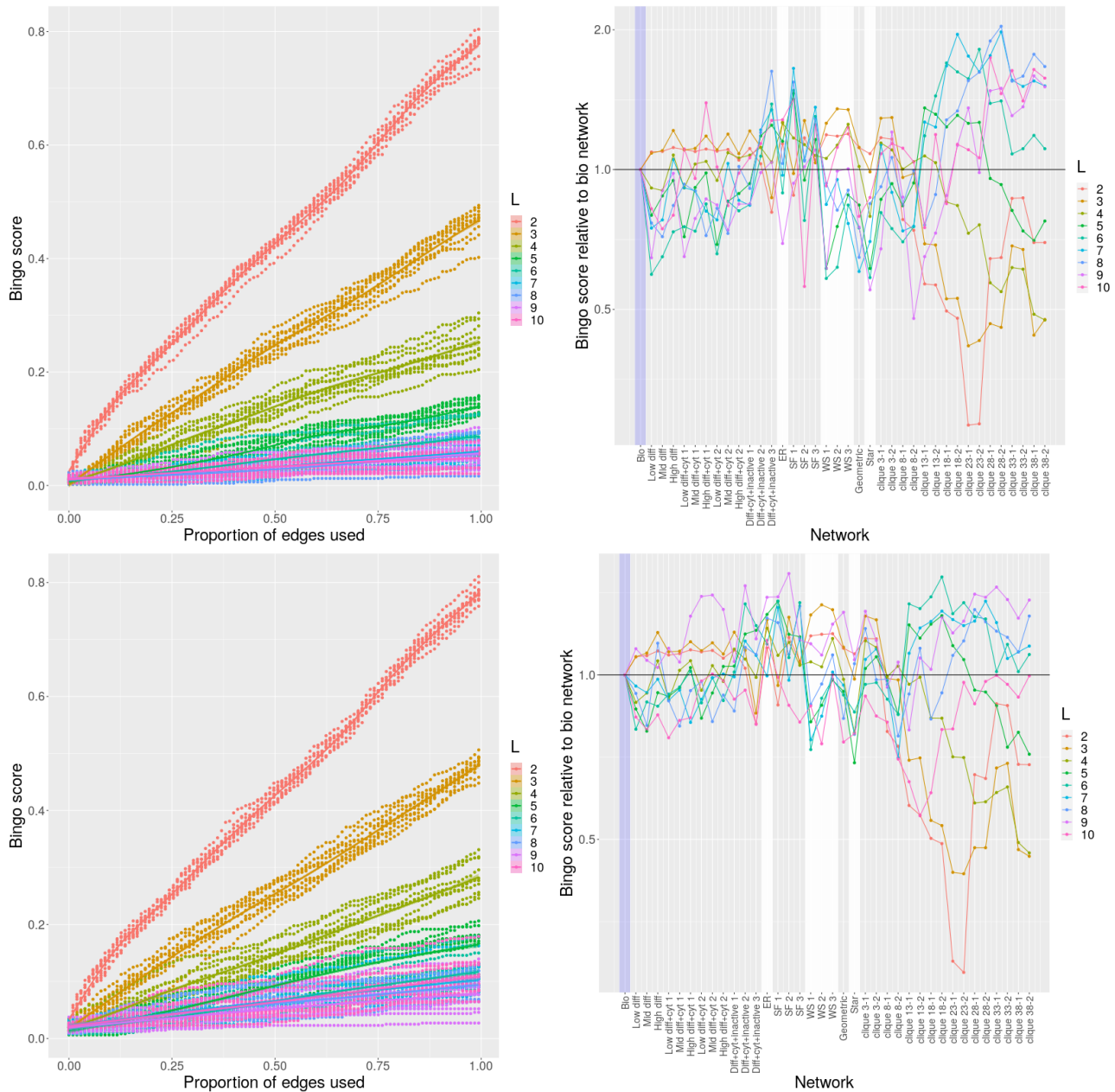
Figure S7: **Effective genome emergence in the presence of master circles.** Bingo score with proportion of edges used for genetic exchanges (analogous to Fig. 2) and final bingo score compared to synthetic networks with matched statistics (analogous to Fig. 3), for different L. (top) a proportion $m = 0.01$ of genetic elements are 'master circles', providing all genetic elements at once; (bottom) $m = 0.02$. Bingo performance is generally increased and more comparable across networks than in the absence of master circles.
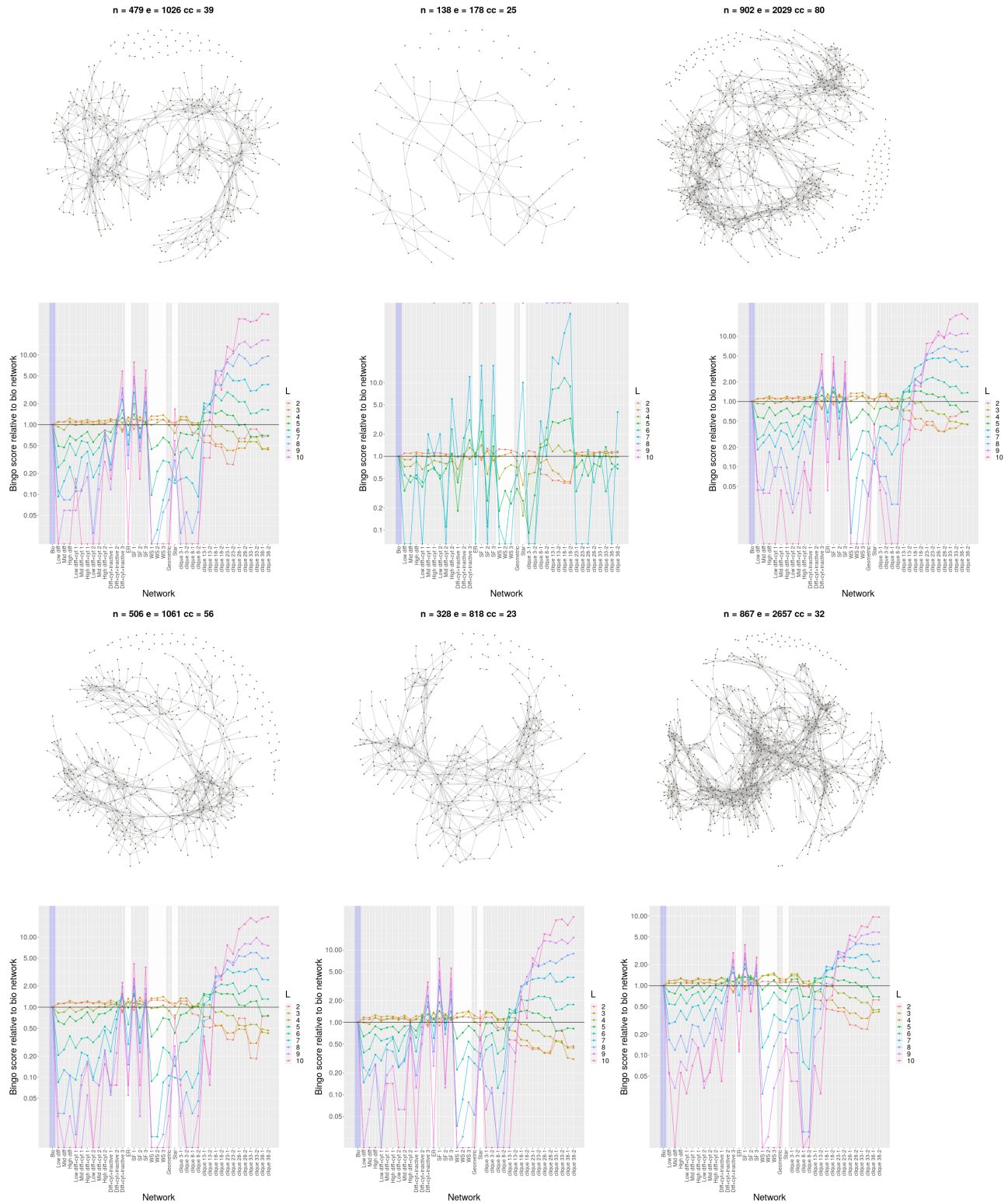
Figure S8: **Comparison of behaviour across different cells.** Examples of encounter network visualisations (*n* nodes, *e* edges, *cc* connected components) and profiles of biological *versus* synthetic partner bingo performance (analogous to Fig. 3) for different single cell observations. All but the top centre cell show very comparable trends. The top centre was unusually small, limiting the size of the mitochondrial population and hence the scale of the encounter network. Correspondingly, the bingo performance for both biological and synthetic partner networks is diminished, especially for high *L*, but the relative performance trends remain comparable.
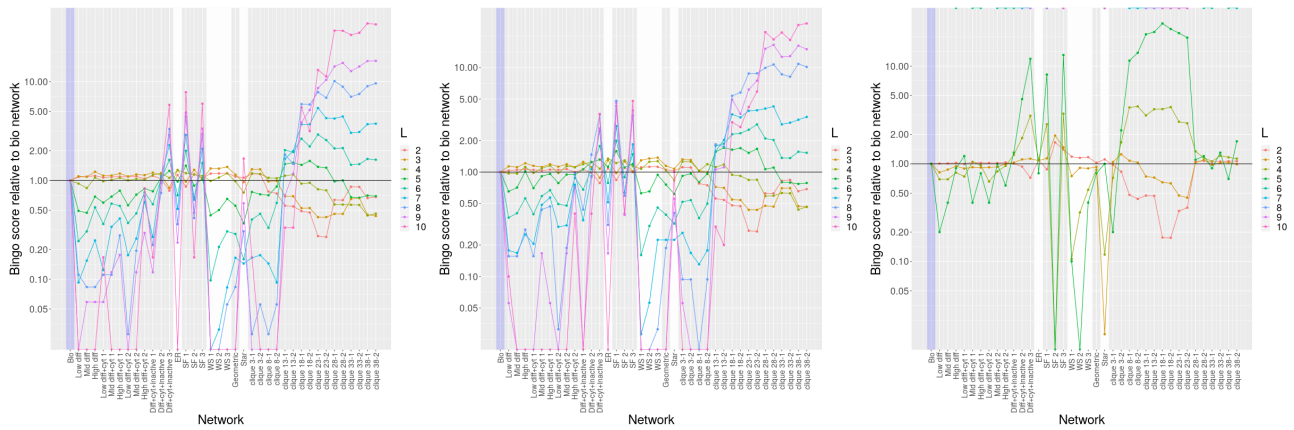
Figure S9: **Comparison of behaviour in different circumstances.** Analogous to Fig. 3, bingo performance for a number of changes to the experimental setup. Left, normal; centre, singletons removed from biological encounter network; right, biological trajectories pruned to a maximum length of ten frames (23 seconds).