The maintenance of multiple ecDNA types in single cancer cells

Elisa Scanu[2], Benjamin Werner[3], Weini Huang[1,2]

[1] Group of Theoretical Biology, Innovation Center for Evolutionary Synthetic Biology, School of Life Science, Sun Yat-sen University, Guangzhou, China

[2] School of Mathematical Sciences, Queen Mary University of London, United Kingdom

[3] Centre for Cancer Genomics and Computational Biology, Barts Cancer Centre, Queen Mary University of London, United Kingdom

Abstract

Extrachromosomal DNA (ecDNA) emerges as a crucial promoter of rapid intra-tumour heterogeneity and oncogene amplification, contributing to tumour progression and adaptability to resistance. Multiple types of ecDNA are often observed within the same cancer cell, promoting high levels of complexity. While the study of multiple ecDNA types is still in its early stages with exciting new clinical and experimental studies in human cancers, few theoretical models have been proposed to understand the mechanisms enabling the maintenance of multiple ecDNA types (mix cells). Here, based on our general framework of modeling multiple ecDNA types, we explore the conditions that allow for a stable coexistence of multiple ecDNA types within single cancer cells. We study how genetic mutations or phenotypic switching among different ecDNA types contribute to the maintenance of ecDNA diversity under different fitness regimes. We show that either positive epistasis or switching is required to stably maintain mix cell populations. Counter-intuitively, intermediate rather than high switching rates lead to the highest fraction of mix cells, which is further confirmed by our analysis of sub-population compositions, transition probabilities between pure and mix states as well as single-cell Shannon indices.

Introduction

Extrachromosomal DNA (ecDNA) has been observed frequently within cancer cells. It allows for highly amplified oncogenes and plays an important role in tumour progression and treatment resistance [3,6,7,18,27,33,36,37,44,52]. Although ecDNA has been observed for many decades in several organisms [51], including human cancer cells [6, 8, 45], the role of ecDNA in cancer promotion and progression only came into focus in recent years, especially as ecDNA-positive tumours tend to have

worse clinical outcome [17, 25]. Indeed, ecDNAs enhance oncogene expression and segregate randomly between daughter cells [2, 23, 25, 33], leading to high cell-to-cell ecDNA copy number variation and rapid evolution [14, 38, 41].

Recent studies found that in about 30% of ecDNAs positive tumours, cells contain multiple ecDNA species carrying different oncogenes or enhancers [30–32]. This coexistence of multiple ecDNAs underscores the complexity of oncogenic processes within cancer cells and facilitates robust inter-ecDNA interactions. For instance, regulatory interactions between ecDNAs, such as enhancer-mediated gene activation, could dramatically enhance the expression levels of oncogenic drivers beyond those achievable by single ecDNA elements [19,30,32], thus conferring a selective advantage to cancer cells [3, 5, 24] and promoting tumour progression [27, 31].

The concept of multiple ecDNA types can be extended beyond ecDNA species carrying distinct oncogenes, through genetic mutations or phenotypic switching. The size of ecDNA elements are typically between 1 to 3 Mbp long [43]. With a mutation rate of 10^{-9} or 10^{-8} per bp per division, this would imply point mutations in the order of 10^{-3} or 10^{-2} per ecDNA copy per division [13, 20]. Previous studies have shown that mutation rates can be even higher because the repair efficiency is often lower on ecDNA compared to chromosomal DNA, resulting in point mutations and small indels upon DNA damage repair [35, 50]. For example, clustered somatic mutations, kyklonic hypermutation, caused by APOBEC3 activity are affecting more than 30% of ecDNAs [16].

Besides point mutations, ecDNA can have higher chromatin accessibility than linear DNA [2, 29, 43, 49], and therefore is highly sensitive to histone modification, such as methylation and acetylation. These are epigenetic mechanisms that can alter chromatine structure and genetic expression without modifying the genetic composition of ecDNAs, which we classify as distinct ecDNA phenotypes. For example, de-methylation, along with copy number amplification, can increase the expression level of certain genes in ecDNA, such as EGFR [1, 14, 47]. On the contrary, methylation can suppress some genes targeted by the human immune system, through for example the KDM5B histone [46], enhancing the epigenetic adaptation of cancer cells. Distinct ecDNA phenotypes can exhibit varying selection strengths influenced by factors such as epigenetic traits.

Understanding the impact of genetic mutations and phenotypic alterations on circular DNA structures has been of significant interest in cancer evolution and disease development. For instance, several studies on mitochondrial DNA (mtDNA) have focused on understanding the impact of methylation [40, 48] and copy number alteration [4, 12, 40], unveiling connections between these factors and the oncogenic features of cells [4,12]. Correspondingly, mathematical models has been developed to investigate the effects of mutations, mtDNA copy number alteration and methylation on population dynamics [21, 26]. However, there are many key biological differences between ecDNA and other circular extra-chromosomal DNA like mtDNA, for exam-

ple the segregation mechanism and cost of losing all circular elements. New models are needed to understand how multiple genotypes or phenotypes of ecDNA elements can coexist in the same cells and population.

We consider a mathematical framework with two main evolutionary parameters: selection, which describes the reproduction rate of a cell based on its ecDNA content, and switching, which represents the probability of either a genetic mutation to occur, in case of multiple ecDNA genotypes, or change in function without genetic alteration, in case of multiple phenotypes [13]. Based on this framework, we investigate how proportions of pure cells carrying a single ecDNA type and mix cells carrying multiple ecDNA types dynamically change under different switching and selection strengths, and the resulting intra-cellular heterogeneity in evolving populations.

Methods

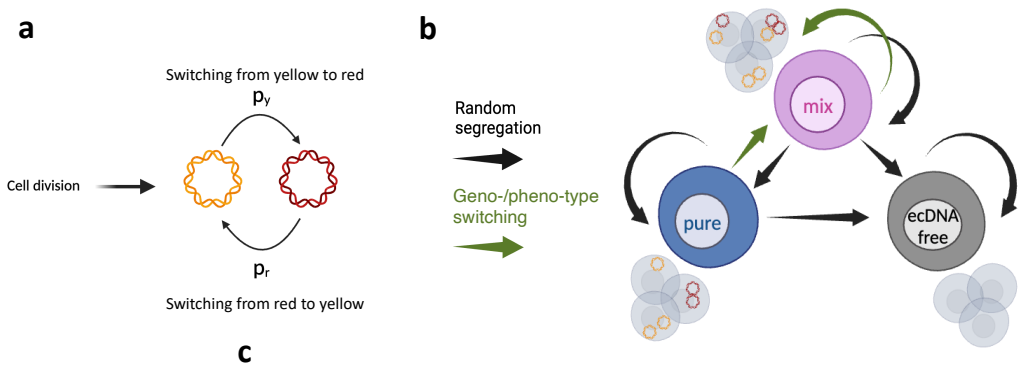
Switching, selection and segregation

We model the dynamics of two ecDNA geno-/pheno-types, represented by respectively the colour yellow and red (Figure 1a) [13]. These two types can change into one another through switching probabilities, p_y and p_r in the range $[0, 1]$. The biological interpretation of these switching probabilities is context-dependent. If we are interested in different genotypes, then switching refers to the probability of obtaining e.g. a point mutation per cell division, which turns one ecDNA copy of the wild genotype into a mutant copy. The switching probabilities in this case are relatively small values in the order of 10^{-3} or 10^{-2} per ecDNA per division. If we consider switching between phenotypes without genetic alteration, slightly larger values of p_y and p_r (but still much smaller than 1) can be interpreted as the probabilities of intrinsic phenotype switching, due to some epigenetic changes such as DNA methylation [14, 28]. While the switching probabilities between distinct ecDNA types can differ in a general framework, here we focus on our analysis in a simplified case where $p_y = p_r = p$. Noteworthy, the special case of $p = 0$ refers to ecDNA species, multiple ecDNA types carrying different oncogene or enhancers are generated in rare events, which does not happen regularly during cell divisions [31].

Another key evolutionary parameter in our model is selection, which is represented by reproduction rates of the yellow and red ecDNA types, denoted by s_y and s_r in the range $[1, +\infty)$ respectively. Cells without ecDNA have a baseline reproduction rate as 1, whilst cells carrying both ecDNA types (mix cells) have a fitness equal to the maximum value between s_y and s_r (Figure 1c).We can relax this assumption and extend our model to other selection scenarios, e.g. imposing additional selective advantage for carrying both ecDNA types [31]. Here, we investigate three different selection regimes: neutral ($s_y = s_r = 1$), identical positive selection ($s_y = s_r > 1$) and non-identical positive selection ($s_y \neq s_r \geq 1$, Figure 1d).

During each cell division, copies of distinct ecDNA types are segregated randomly

and independently into daughter cells, mathematically modeled by a binomial distribution with probability $\frac{1}{2}$. Again we can relax this assumption and introduce non-random co-segregation of distinct ecDNA types [31]. Here, we focus on a baseline scenario where there is no co-segregation between different ecDNA geno-/pheno-types.



c

Selection scenario	Selection coefficient value			
	ecDNA-free	pure yellow	pure red	mix
Neutral selection	1	1	1	1
Identical positive selection	1	s (>1)	s (>1)	s (>1)
Non-Identical selection	1	s_y (>1)	1	s_y (>1)

Figure 1: Switching and selection. **a. Switching.** At each cell division, each ecDNA can change colour with probabilities p_y and p_r. **b. Transition graph of different ecDNA subpopulations.** Blue represents cells carrying one ecDNA type, i.e. pure; purple represents cells carrying both yellow and red elements, i.e. mix; grey represents cells without ecDNA, i.e. ecDNA-free. The diagram shows the mutual dependency for subpopulations, where the black arrows track the relations in case of simple random segregation and the green arrows add the additional relations with the incorporation of type switching. **c. Selection scenarios.** Quantitative description of the different selection scenarios investigated throughout the paper.

Classification of subpopulations

We define four different subpopulations based on their ecDNA content: pure yellow, pure red, mix and ecDNA-free cells. A cell is classified as pure if it carries just one ecDNA geno-/pheno-type, either yellow or red, whilst a cell is defined as mix if it carries both types (Figure 1b). At each cell division, a mother cell may generate daughters belonging to different subpupulations, based on the interplay of random ecDNA segregation and switching (Figure 1b). We assume that ecDNA initiation is rare, thus a cell losing all ecDNA copies does not gain a new copy spontaneously in

divisions and its offsprings remain ecDNA-free.

Results

Starting with a single copy of the yellow ecDNA type in the initial cell and the same probability of switching for both ecDNA types, i.e. $p_y = p_r = p$, we vary switching probabilities and selection regimes to understand their implications on the fraction and ecDNA copy number distribution of different subpopulations.

EcDNA copy number distributions among subpopulations are diversified by selection and switching

When starting with a single cell that carries 1 yellow ecDNA copy, pure yellow cells dominate the population and have a higher copy number distribution than pure red cells in all selection scenarios for small switching probabilities (Figure 2a). We observe a clear difference among neutral, identical and non-identical positive selection, where the copy number distributions of pure cells are highest under non-identical selection, followed by the ones under identical selection, then by those under neutral selection. The copy number distributions of mix cells are similar among different selection scenarios probably due to their low abundance. In contrast, when the switching probability between yellow and red ecDNA elements is large (Figure 2c), the copy number distributions between pure yellow and red cells are identical for all selection scenarios, although pure yellow cells have higher fitness than pure red cells under non-identical positive selection. This means that selection plays little role in the dynamics of pure cells when the switching probability is extremely high.

The copy number distribution of mix cells is quite similar for small and large switching probabilities (see purple lines in Figure 2). This changes when switching probabilities are intermediate (Figure 2b). The distribution of mix cells (purple lines) is higher compared to small and large switching probabilities for all selection regimes (purple dashed and dash-dotted lines). The distributions of yellow and red pure cells are narrow and peaked at low copy numbers with very short tails (yellow and red lines), highlighting the predominance of mix cells under intermediate switching probabilities.

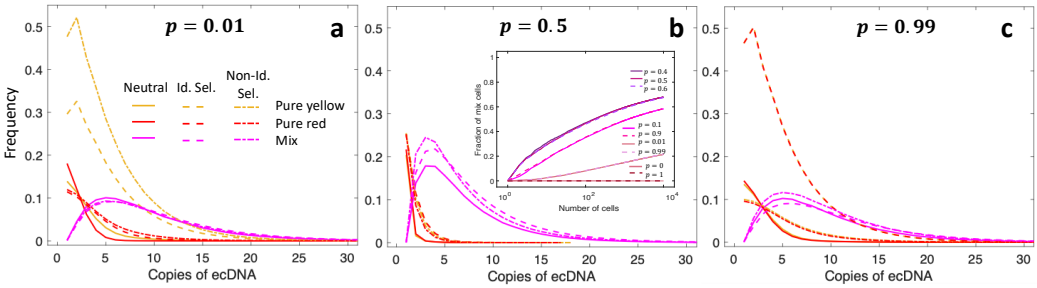


Figure 2: **Distribution of ecDNA copies among different subpopulations**. We show the ecDNA copy number distributions for pure yellow, pure red and mix ecDNA cells under small, intermediate and large switching probabilities when $p_y = p_r = p$. For each switching choice, we present results under three selection scenarios defined in Figure 1c with $s = 2$. That is under neutral selection, all cells have fitness as 1; under identical selection, all cells carrying ecDNA have fitness as 2; under non-identical selection, the pure yellow and mix cells have fitness as 2, while the pure red cells have fitness as 1 as ecDNA-free cells. The internal panel in **b** shows the fraction of mix ecDNA cells under a broad range of switching probabilities (p between 0 and 1). We start our simulations from a single cell with 1 yellow ecDNA copy, and show the distribution recorded at 10^5 cells over 1000 realisations.

The fraction of mix ecDNA cells is highest under intermediate switching

Intuitively, we would expect that higher switching probabilities will promote higher fractions of mix ecDNA cells. However, our results show a contrary pattern, where intermediate switching probabilities lead to the highest fraction of mix cells for all selection regimes (Figure 2b inner panel). We first suspected that this might be related to our initial condition, where the first cell carries only a single copy of yellow ecDNA. Thus, high switching probabilities could lead to synchronous switching of all copies in cells between yellow and red types initially, which dominates the dynamics later.

To exclude the impact of this initial condition, we vary the starting time of switching under high switching probabilities (Figure 3). When the population grows, there is a fast accumulation of copy number diversity among cells. Switching early or later will naturally capture different initial states of copy number distributions where the switching starts to act. Our results show a convergence of the fractions of sub-populations from different switching times under all selection regimes. This convergence happens faster under neutral selection (Figure 3a) than identical (Figure 3b) and non-identical positive selection (Figure 3c). This indicates that the fractions of sub-populations are independent of the initial copy number distribution of the population. The observation of the highest fractions of mix cells under intermediate switching probabilities is not a consequence of the initial switching condition.

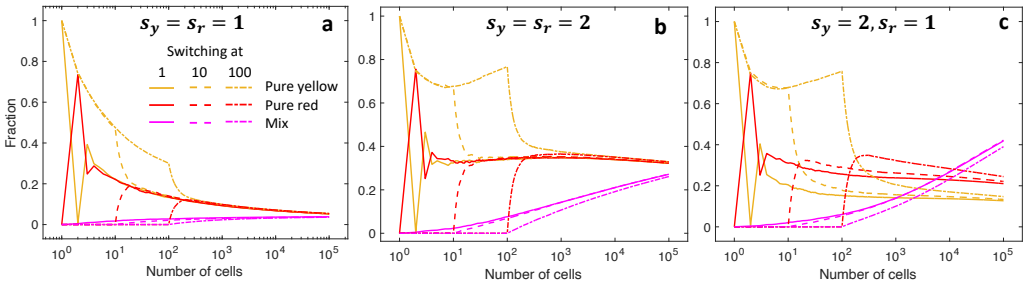


Figure 3: **Fraction of subpopulations over time for different starting points of switching.** **a.** Neutral selection. **b.** Identical positive selection. **c.** Non-identical selection. In each panel, we compare the fractions of pure and mix ecDNA subpopulations, when the switching probability is high ($p_y = p_r = p = 0.99$). The start of switching is introduced when the population grows from a single cell to 1, 10 and 100 cells. The initial cell carries 1 yellow ecDNA copy. All dynamics are averaged over 1000 realisations.

Switching and selection are both required for maintaining mix ecDNA cell populations

We further investigate how and when mix cells can be stably maintained in growing populations. Through the transition diagram between subpopulations (Figure 1b), we see that the ecDNA-free state is an absorbing state under neutral selection with or without switching. Indeed, we observe that ecDNA-positive cells (Figure 4a solid coloured lines) including the subpopulation of mix cells (solid purple line) rapidly diminish during the population expansion when selection is neutral. Under the same switching probability ($p = 0.1$), when we add selection for ecDNA-positive cells, the fraction of mix cells (Figure 4a dashed purple line) increases over time.

However, when mix cells have the same positive selection as pure cells ($s_y = s_r = s = 2$) but switching is off ($p = 0$) between the two ecDNA types, mix cells will get lost over time (Figure 4b solid purple line). This corresponds to removing the green arrows related to switching in Figure 1b, which makes the pure cells an absorbing state under positive identical selection. Only the simultaneous action of positive selection and active switching ensures the maintenance of the mix subpopulation (Figure 4a&b, dashed purple lines) in the absence of co-selection of multiple ecDNA species.

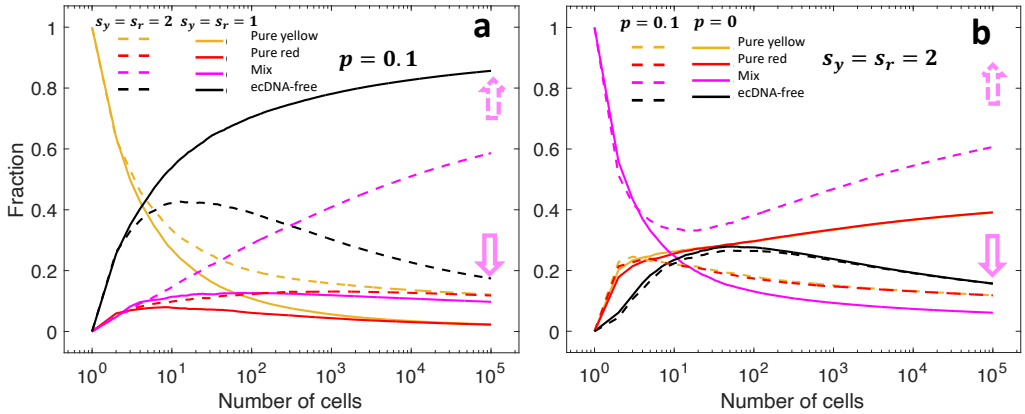


Figure 4: **Mix cells are maintained if selection is positive for cells carrying ecDNA and switching is on.** We compare temporal dynamics of the fractions of pure, mix and ecDNA-free cells when selection and switching are on and off. **a.** We fix the switching probability $p = 0.1$, mix cells are maintained over time only if selection is positive ($s_y = s_r = 2$) emphasised by the purple arrows. The initial condition for simulations in this panel is a single cell with 1 yellow ecDNA copy. **b.** We fix the selection strength to be positive and identical ($s_y = s_r = 2$), and compare dynamics when switching is on or off. Again following the purple arrows, we can see the fraction of mix cells only increases if switching is on. The initial condition is a single mix cell with 1 yellow and 1 red ecDNA copy. All dynamics are averaged over 1000 realisations.

Transition between different subpopulations during cell divisions

We further investigate this dynamics by measuring the probabilities of transitioning between different subpopulations, i.e. pure, mix, ecDNA-free states, during cell divisions in simulations. We let the population grow to 10^5 cells and record the frequency of transitions from the state of the mother cell into possible states of both daughter cells (Figures S4-S6) for each division, thus quantifying all arrows in Figure 1b. We also record frequencies of daughter cells to remain in a mix, pure or ecDNA-free state regardless of the nature of their mothers (Figures 5 and S3-S6), thus quantifying the probabilities of staying in each state in Figure 1b. This allows us to further understand the role of switching probabilities in determining sub-population compositions.

Considering identical positive fitness for both ecDNA types ($s_y = s_r = s$), the probability of being ecDNA-free is independent of the switching probability as expected (Figures 5c and S3c). During the population expansion, the mean ecDNA copy number increases when selection for cells carrying ecDNA is positive. The increasing mean copy agrees with the expected observation that the probabilities of losing ecDNA during cell divisions and thus staying in the ecDNA-free state decrease

with increasing population size (Figures 5c and S3c). On the contrary, the probability of staying in the mix ecDNA state increases over time and it peaks when p is intermediate (Figures 5a and S3a and Figure 2b inner panel). Correspondingly, we observe the opposite behaviour for pure ecDNA cells (Figures 5b and S3b). The probability of staying in the pure state (either pure yellow or pure blue cells) has the lowest value under intermediate switching probabilities.

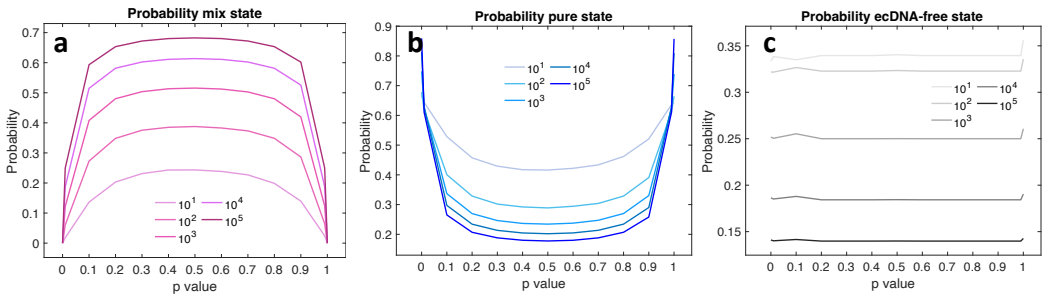


Figure 5: **Probability of staying in mix, pure and ecDNA-free states at different times (population size) under positive identical selection ($p_y = p_r = p$, initial condition: 1 single pure cell with 1 yellow copy).** **a.** Probability of staying in mix state. **b.** Probability of staying in pure state **c.** Probability of staying in ecDNA-free state. These dynamics are generated by simulations starting with a single cell with 1 yellow ecDNA copy and results are recorded over 1000 realisations.

Interestingly, for low (between 0 and 0.05) or high (between 0.95 and 1) switching probabilities, the probabilities to remain in pure state (Figures 5b and S3b) have opposite order compared to scenarios of intermediate switching. We see that in two limiting cases, i.e. ecDNA elements never ($p = 0$) or always ($p = 1$) switch between different types, without additional fitness advantages mix cells will be lost after the first few divisions and the probability of being in pure states approaches 1. The dynamics of these two extreme cases must be symmetrical, which is confirmed by our simulations starting with different ecDNA configurations in the initial cell (Figure S3).

For intermediate switching rates, transitions from pure into mix cells are more likely than from mix into pure cells (Figure S4 d-f), explaining the higher prevalence of mix cells in that parameter regime. If we vary selection strength ($s = 2$ compared to $s = 3.5$), the probabilities of staying in different sub-populations states change as expected, see Figure S4 black ($s = 2$) and red numbers ($s = 3.5$) inside the circles. A higher selective advantage leads to higher probabilities of staying in ecDNA-positive states (pure and mix) and a lower probability of being ecDNA-free. However, the selection strength has a mild impact on the transition probabilities between states.

For neutral selection ($s = 1$), the probability of staying in the mix state decreases considerably (Figure S5), and mix cells will get lost over time. Again, the transition

probabilities (arrows in the transition diagrams) only deviate marginally from the identical positive selection regime (Figure S4). This further confirms that, without additional fitness advantage for mix cells (identical positive or neutral selection), random segregation (black arrows) and switching (green arrows) are the main forces that dominate the transition between different sub-populations rather than selection strength. However, selection drives the reproduction difference between ecDNA-positive and ecDNA-free cells, and thus is also required for the maintenance of mix cells. Without selection, pure cells upon which switching acts are less abundant, and there is no source to compensate the constant loss of mix cells due to random segregation.

All these observations remain qualitatively true if we consider different initial conditions, e.g. when the population starts from a single mix cell of one yellow and one red copy. The highest percentage of mix cells is still observed under intermediate p, with a small quantitative change of transition probabilities and probabilities of staying in different subpopulations (Figure S6).

Shannon diversity at single-cell level justifies distributional results

We have studied the composition and copy number distributions of sub-populations and demonstrated how they are impacted by selection and switching. We further explore the diversity in the population by using the Shannon indexes defined at the whole population and single-cell level given by

$$H = -\sum_k c_k \ln c_k, \tag{1}$$

where c_k represents the density of category k in the population (Table 1).

Shannon index	**Number of categories**	**Categories**
whole population	dynamical, [1,+∞]	$\rho_{i,j}$: densities of cells with i yellow and j red ecDNA copies
single-cell	2	density of yellow ecDNA copies and density of red ecDNA copies in each single cell

Table 1: Categories in Shannon index. For the Shannon index at the whole population level, every configuration of copies classified by i yellow and j red copies, is a category, and the number of categories is dynamical ranging from 1 to $+\infty$. For the Shannon index at single-cell level, there are two categories, red and yellow elements.

We first looked at the Shannon index at the whole population level for different switching probabilities under neutral, identical and non-identical positive selection (Figure S1). We see a clear difference between neutral and positive selection as expected [25]. Under neutral selection, we observe a cluster of low mean ecDNA copy number per cell and small Shannon index values (Figure S1 light grey circles).

When selection is positive, the mean ecDNA copy number per cell increases with the fraction of ecDNA-positive cells over time (Figures 3 and 4), reflected by an increasing Shannon index (Figure S1 dark grey and black circles). On the contrary, we hardly observe any distinguishable difference in the Shannon index values across different switching probabilities for all selection scenarios. However, in our analysis above, both selection and switching are required for maintaining mix cells, and can impact the ecDNA copy number distribution of subpopulations. We conclude that the impact of switching is not reflected in the Shannon index at the whole population level.

We then moved to the Shannon index at a single-cell level and look at the diversity of the yellow and red ecDNA copies within mix cells. Starting from a mix cell carrying 1 yellow and 1 red ecDNA copies, we record the Shannon index values of all mix cells over time and present their distributions at a given population size. Showing the single-cell Shannon index over its corresponding ecDNA copy number, we have a discrete distribution of all possible values of Shannon index for any given copy number (Figure 6 grey dots). For example, for a mix cell with 2 ecDNA copies, there is only one possible Shannon index

$$-\sum_{k=1}^{2} \frac{1}{2} \ln \frac{1}{2},$$

which is around 0.695. This is also the highest Shannon index value a single cell can have independent of its copy number, because the Shannon index decreases when the ratio of yellow and red cells is uneven (Figure 6 horizontal line and Figure S2). In Figure 6, if we join the lowest values of possible Shannon index across all copy numbers, this discrete curve represents all mix cells with the lowest diversity by having only one yellow or red copy. While their Shannon index will be

$$-(\frac{1}{\text{Copy number}} \ln \frac{1}{\text{Copy number}} + (1 - \frac{1}{\text{Copy number}}) \ln (1 - \frac{1}{\text{Copy number}})),$$

we see a decreasing curve over copy number. A similar logic applies to other discrete curves composed of the grey dots, which refer to the cells with the 2nd, 3rd lowest diversity and so on. Such mix cells with a low Shannon diversity are more likely to generate pure daughter cells in random segregation, due to the imbalance in copy numbers between the yellow and red ecDNA elements in those cells.

We show the occupation of those possible Shannon index values in stochastic simulations from no switching to high switching probability (Figure 6 coloured dots). Different from the Shannon index at the whole population level, the overall configuration of single-cell Shannon index is significantly influenced by switching. The occupation of low Shannon index curves decreases when the switching probability p increases from small to intermediate values (Figure 6 b to a), but increases again when p becomes very large (Figure 6 c to d). This demonstrates that the dynamics of

mix cells with low Shannon index values indeed contribute to the transition between mix and pure cells. Furthermore, we observe an invariant peak in the histogram of Shannon index at its highest possible value (Figure 6 and S2), corresponding to configurations with equal numbers of yellow and red copies in those cells. This suggests a consistent proliferation behaviour that tends to stabilize toward a more homogeneous configuration between the two ecDNA types in mix cells, regardless of the values of the switching probability.

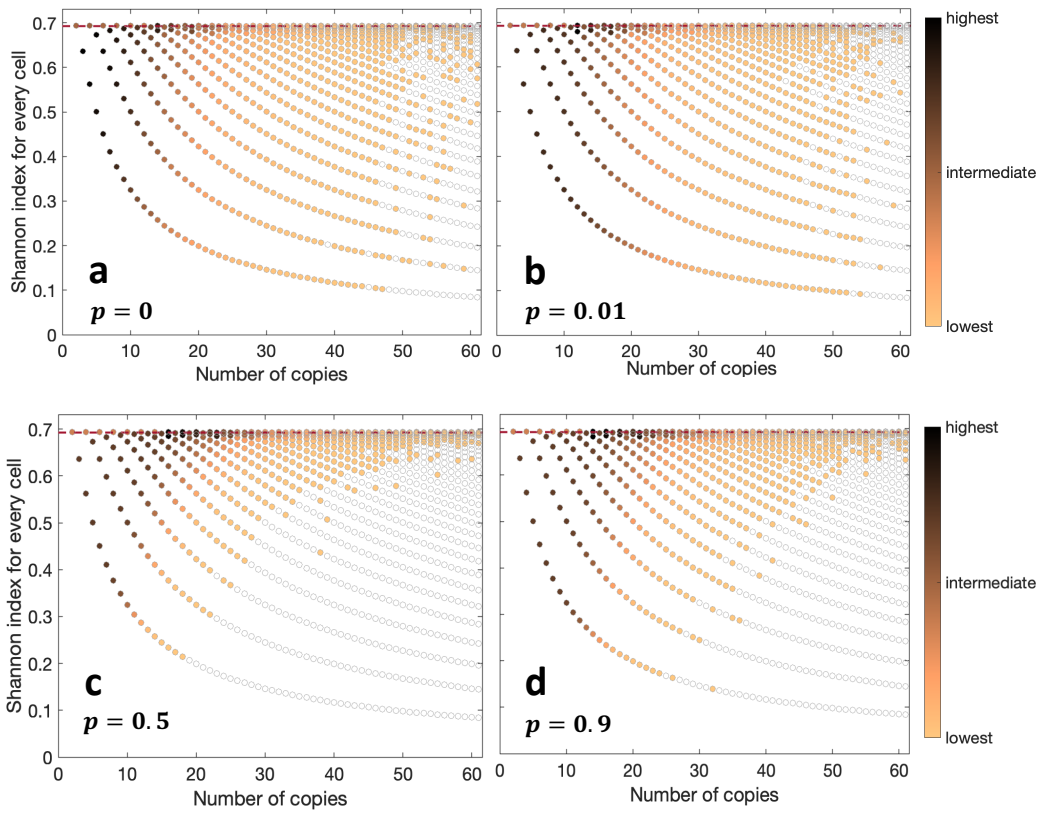


Figure 6: **Shannon index at single-cell level. Scatter plots of Shannon index values for four different values of** p **(**$s_y = s_r = s = 2$**).** Each value of the Shannon index is coloured based on its frequency in the population. The horizontal line corresponds to the most frequent Shannon index value. The grey dots on the background are all the possible values on the Shannon index, regardless of p. No switching (**a**) versus low (**b**), intermediate (**c**) and high (**d**) switching scenarios are compared. The initial configuration for these simulations is a single cell with 1 yellow and 1 red copy. The indices are recorded at a population size of 10^4 cells over 600 realisations.

Discussion

We have built a general framework to model ecDNA species with and without geno-/pheno-types switching [13]. Based on this framework, here we further explore the mechanisms necessary for maintaining multiple ecDNA geno-/pheno-types in growing tumours. We focus on a model with two ecDNA types, yellow and red and distinguish pure cells with just one type of ecDNA, mix cells with two types of ecDNA, and ecDNA free cells. We develop a stochastic model of multiple ecDNA types with two key parameters, the selection coefficient and geno-/pheno-type switching. Selection can be neutral for the entire population, or positive and identical for cells carrying ecDNA, or non-identical between cells carrying yellow or red type of ecDNA. Switching can be interpreted as a genetic change (with low probability) or a phenotypic change (with a higher probability) during cell division [14].

Variation of these parameters profoundly affects the distribution of ecDNA copy numbers within cell subpopulations (Figure 2). In particular, the switching probability p significantly influences copy number distributions. Our findings suggest that values of p close to 0.5 broaden the distribution of ecDNA copy numbers in mix cells. Intermediate switching enhances the maintenance of both ecDNAs within cells (Figure 2b, internal panel). Long-term coexistence of multiple geno-/pheno-types is achievable only if ecDNA positive cells have a selective advantage and switching is possible (Figure 4). The absence of either condition leads to the extinction of mix cells ultimately. It is noteworthy that distinct ecDNA species can be maintained in tumours without active switching, if mix cells have additional fitness advantage than pure cells (named as co-selection in [13, 31]). This for example refers to co-amplification of different oncogenes on multiple ecDNA species.

The observation that intermediate p values result in higher fractions of mix cells is reflected by the state transition probabilities (Figures 5 and S3). When switching is intermediate, the probabilities of transitioning from pure to mix states increases and the probabilities of transitioning from mix to pure states decreases (Figures S4-S6). While switching rather than selection plays a dominating role in those transition probabilities, the incorporation of positive selection allows the increase of ecDNA-positive cells (Figure S5 versus S4 and S6), on which switching can act. Thus, both selection and switching are required to maintain mix cells without co-selection.

We measured the intra-cellular heterogeneity by the single-cell Shannon index. Focusing on the mix-cell subpopulation, the probability density of cells with low Shannon index decreases and increases again for low and high switching rates (Figure 6). This shows imbalanced copy numbers of yellow and red ecDNA elements in majority of mix cells under low or high switching rates. These mix cells are more likely to generate pure cells under the random segregation of ecDNA copies during cell divisions. This further explains why we observe higher transitioning probabilities of mix to pure state under low and high switching rates, as well as a higher fraction

of mix cells for intermediate switching rates.

Models of selection and maintenance of multiple ecDNA types are of immediate interest in cancer research. About 30% of all human cancers contain ecDNA under strong positive selection [33]. Although the intra-cellular variation of ecDNAs, especially mutational and epimutational profiles leading to multipe ecDNA types are not yet well-understood, they are likely to further contribute to the fast evolvability of ecDNA-positive tumours [7, 14]. By analysing necessary mechanisms to maintain variation of mutational or epimutational processes on ecDNAs within cells, our model can apply to understand complex ecDNA dynamics under genetic mutations and phenotypical changes.

Acknowledgements

E.S. is supported by the Engineering and Physical Sciences Research Council (EPSRC) PhD fellowship that is part of the United Kingdom Research and Innovation association (UKRI). B.W. is supported by a Barts Charity Lectureship (grant no. MGU045) and a UKRI Future Leaders Fellowship (grant no.MR/V02342X/1). W.H.is funded by NNSF General Program (grant no. 3217024). E.S. and W.H. are funded by Cancer Grand Challenges CGCSDF-2021\100007 with support from Cancer Research UK and the National Cancer Institute.

References

[1] Morton AR, Dogan-Artun N, Faber ZJ, et al. Functional enhancers shape extrachromosomal oncogene amplifications. *Cell*, 179(6):1330–1341, 2019.

[2] Bailey C, Shoura MJ, Mischel PS, and Swanton C. Extrachromosomal DNA-relieving heredity constraints, accelerating tumour evolution. *Annals of Oncology*, 31:884–893, 2020.

[3] Aktipis CA, Boddy AM, Gatenby RA, Brown JS, and Maley CC. Life history trade-offs in cancer evolution. *Nat. Rev. Cancer.*, 13:883–892, 2013.

[4] Castellani CA et al. Thinking outside the nucleus: Mitochondrial DNA copy number in health and disease. *Mitochondrion*, 53:214–223, 2020.

[5] Storlazzi CT, Lonoce A, Guastadisegni MC, Trombetta D, et al. Gene amplification as double minutes or homogeneously staining regions in solid tumors: origin and structure. *Genome Research*, 20(9):1198–1206, 2010.

[6] Cox D, Yuncken C, and Spriggs AI. Minute chromatin bodies in malignant tumours of childhood. *Lancet*, 1:55–58, 1965.

[7] Nathanson DA et al. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science*, 343:72–76, 2014.

[8] Von Hoff DD et al. Amplified human MYC oncogenes localized to replicating submicroscopic circular DNA molecules. *Proceedings of the National Academy of Sciences*, 85(13):4804–4808, 1988.

[9] Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[10] Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[11] Gillespie DT. Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, 58:35–55, 2007.

[12] Reznik E et al. Mitochondrial DNA copy number variation across human cancers. *eLife*, 5:e10769, 2016.

[13] Scanu E, Werner B, and Huang W. Population dynamics of multiple ecDNA types. Unpublished manuscript available at: `https://github.com/eliscanu96/popdynamics/blob/main/Population_dynamics_of_multiple_ecDNA_types.pdf`, 2024.

[14] Yi E, Chamorro González R, Henssen AG, et al. Extrachromosomal DNA amplifications in cancer.105. *Nature Reviews Genetics*, 23:760–771, 2022.

[15] Pielou EC. The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology*, 13:131–144, 1966.

[16] Bergstrom EN et al. Mapping clustered mutations in cancer reveals APOBEC3 mutagenesis of ecDNA. *Nature*, 602:510–517, 2022.

[17] Kim H et al. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat. Gen.*, 52:891–897, 2020.

[18] Møller HD, Mohiyuddin M, Prada-Luengo I, Sailani MR, Halling JF, Plomgaard P, et al. Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nature Communications*, 9(1):1069, 2018.

[19] Møller HD, Mohiyuddin M, Prada-Luengo I, Sailani MR, Halling JF, Plomgaard P, et al. CRISPR-c: circularization of genes and chromosome by CRISPR in human cells. *Nucleic Acids Research*, 46(22):e131, 2018.

[20] Martincorena I and Peter JC. Somatic mutation in cancer and normal cells. *Science*, 349.625:1483–1489, 2015.

[21] Johnston IG and Jones NS. Evolution of cell-to-cell variability in stochastic, controlled, heteroplasmic mtDNA populations. *The American Journal of Human Genetics*, 99(5):1150–1162, 2016.

[22] Ludwig JA and Reynolds JF. *Statistical ecology: a primer in methods and computing.*, volume 1. John Wiley and Sons, 1988.

[23] Rose JC, Belk JA, Wong ITL, Luebeck J, HT Horn, Daniel B, Jones MJ, Yost KE, Hung KL, Kolahi KS, Curtis EJ, Kuo CJ, Bafna V, Mischel PS, and Chang HY. Disparate pathways for extrachromosomal DNA biogenesis and genomic DNA repair. *Cancer Discovery*, 2024.

[24] Cowell JK. Double minutes and homogeneously staining regions: gene amplification in mammalian cells. *Annual review of genetics*, 16(1):21–59, 1982.

[25] Lange JT, Rose JC, Chen CY, et al. The evolutionary dynamics of extrachromosomal DNA in human cancers. *Nature Genetics*, 54:1527–1533, 2022.

[26] Kornick K, Bogner B, Sutter L, and Das M. Population dynamics of mitochondria in cells: A minimal mathematical model. *Frontiers in Physics*, 7:146, 2019.

[27] Song K et al. Plasticity of extrachromosomal and intrachromosomal BRAF amplifications in overcoming targeted therapy dosage challenges. *Cancer Discovery*, 12:1046–1069, 2022.

[28] Johnson KC et al. Single-cell multimodal glioma analyses identify epigenetic regulators of cellular plasticity and environmental stress response. *Nature Genetics*, 53:1456–1468, 2021.

[29] Hung KL, Luebeck J, Dehkordi SR, et al. Targeted profiling of human extrachromosomal DNA by CRISPR-CATCH. *Nat. Genet.*, 54:1746–1754, 2022.

[30] Hung KL, Yost KE, Xie L, et al. ecDNA hubs drive cooperative intermolecular oncogene expression. *Nature*, 600:731–736, 2021.

[31] Hung KL, Jones MG, Wong IT, et al. Coordinated inheritance of extrachromosomal DNA species in human cancer cells. *bioRxiv*, 2023.

[32] Hung KL, Mischel PS, and Chang HY. Gene regulation on extrachromosomal DNA. *Nature Structural and Molecular Biology*, 29:736–744, 2022.

[33] Turner KM, Deshpande V, Beyter D, Koga T, Rusert J, Lee C, Li B, Arden K, Ren B, Nathanson DA, et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature*, 543:122–125, 2017.

[34] Jost L. Entropy and diversity. *Oikos*, 113:363–375, 2006.

[35] Zuberi L, Adeyinka A, and Kuriakose P. Rapid response to induction in a case of acute promyelocytic leukemia with MYC amplification on double minutes at diagnosis. *Cancer Genetics and Cytogenetics*, 198(2):170–172, 2010.

[36] Pecorino LT, Verhaak RGW, Henssen A, and Mischel PS. Extrachromosomal DNA (ecDNA): an origin of tumor heterogeneity, genomic remodeling, and drug resistance. *Biochemical Society Transactions*, 50(6):1911–1920, 2022.

[37] Andriy M, Janiszewska M, and Polyak K. Intratumor heterogeneity: the rosetta stone of therapy resistance. *Cancer cell*, 37(4):471–484, 2020.

[38] McGranahan N and Swanton C. Clonal heterogeneity and tumor evolution: Past, present, and the future. *Cell.*, 168(4):613–628, 2017.

[39] Gotelli NJ and Colwell RK. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, 4:379–391, 2001.

[40] Kelly RD, Mahmud A, McKenzie M, Trounce IA, and St John JC. Mitochondrial DNA copy number is regulated in a tissue specific manner by DNA methylation of the nuclear-encoded DNA polymerase gamma A. *Nucleic Acids Research*, 40(20):10124–10138, 2012.

[41] Verhaak RG, Bafna V, and Mischel PS. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nature Reviews Cancer*, 19:283–288, 2019.

[42] Nakaoka S and Aihara K. Stochastic simulation of structured skin cell population dynamics. *Journal of Mathematical Biology*, 66:807–835, 2013.

[43] Wu S et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature*, 575:699–703, 2019.

[44] Wu S, Bafna V, and Mischel PS. Extrachromosomal DNA (ecDNA) in cancer pathogenesis. *Current Opinion in Genetics and Development*, 66:78–82, 2021.

[45] Carroll SM, DeRose ML, Gaudray P, Moor CM, Needham-Vandevanter DR, Von Hoff DD, and Wahl GM. Double minute chromosomes can be produced from precursors derived from a chromosomal deletion. *Molecular and cellular biology*, 1988.

[46] Zhang SM, Cai WL, Liu X, Thakral D, Luo J, Chan LH, et al. KDM5B promotes immune evasion by recruiting SETDB1 to silence retroelements. *Nature*, 598:682–687, 2021.

[47] Clarke TL, Tang R, Chakraborty D, Van Rechem C, Ji F, Mishra S, et al. Histone lysine methylation dynamics control EGFR DNA copy-number amplification. *Cancer Discovery*, 10:306–325, 2020.

[48] Patil V, Cuenin C, Chung F, Aguilera JRR, Fernandez-Jimenez N, Romero-Garmendia I, et al. Human mitochondrial DNA is extensively methylated in a non-CpG context. *Nucleic Acids Research*, 47(19):10072–10085, 2019.

[49] Chen W et al. Sequencing of methylase-accessible regions in integral circular extrachromosomal dna reveals differences in chromatin structure. *Epigenetics and Chromatin*, 14:1–11, 2021.

[50] Dong Y, He Q, Chen X, Yang F, He L, and Zheng Y. Extrachromosomal DNA (ecDNA) in cancer: mechanisms, functions, and clinical implications. *Frontiers in Oncology*, 13, 2023.

[51] Hotta Y and Alix B. Molecular size and circularity of DNA in cells of mammals and higher plants. *Proceedings of the National Academy of Sciences*, 53(2):356–362, 1965.

[52] Liao Z, Jiang W, Ye L, Li T, Yu X, and Liu L. Classification of extrachromosomal circular DNA with a focus on the role of extrachromosomal DNA (ecDNA) in tumor heterogeneity and progression. *Biochimica et Biophysica Acta Reviews on Cancer*, 1874(1), 2020.

Supplementary

Stochastic simulations

The simulations presented in this paper are conducted through MATLAB (version 2022). Our approach is based on the Gillespie Algorithm [9–11, 42]. In each simulation iteration, the precise count of ecDNA copies is recorded. Cells are chosen randomly but proportional to fitness for proliferation. Cells without ecDNA are assigned a proliferation rate (s) of 1, with time measured in generations. The introduction of a fitness effect for cells carrying ecDNA is represented by a proliferation rate $s > 1$, indicating a fitness advantage, while $0 < s < 1$ denoted a fitness disadvantage, and $s = 1$ signified no discernible fitness difference. Upon selection for proliferation, cells with ecDNA double their ecDNA copies, which are randomly distributed to both daughter cells through a binomial trial with a success rate of $\frac{1}{2}$. Conversely, cells without ecDNA do not pass ecDNA to their daughters. Simulations are terminated upon reaching a specified cell population size, which in the majority of the experiments is either 10^4 or 10^5. The recorded output includes the ecDNA copy number for each cell at the conclusion of each simulation, facilitating the derivation of various analytical metrics such as the ecDNA copy distribution, deterministic dynamics, stochastic dynamics, and Shannon diversity indices.

Entropy and evenness of the system through Shannon diversity indices

The Shannon diversity index, often referred to simply as Shannon entropy, represents a fundamental measure of the entropy within a system. In the context of ecological studies and population dynamics, the Shannon diversity index provides

insights into the diversity, evenness, and distribution of different entities within a population [15, 22, 34, 39]. Specifically, when applied to the distribution of ecDNA copies among cells, the Shannon diversity index offers a quantitative assessment of the variability and uniformity in ecDNA abundance across the cellular population. The formula for this index is:

$$H = -\sum_k c_k \ln c_k, \tag{S1}$$

where c_k represents the density of category k in the population. By summing the products of each category's density and the natural logarithm of its density, the Shannon index captures the overall uncertainty or information content associated with the distribution of categories.

It is crucial to note that the Shannon diversity index is sensitive to the relative abundances of different categories within the population. When the abundances of ecDNA types are uneven, with certain variants being more prevalent than others, the resulting Shannon entropy value decreases. Conversely, a more equitable distribution of ecDNA copies among cells leads to a higher Shannon entropy value, indicating greater diversity and evenness in the population.

Supplementary Figures

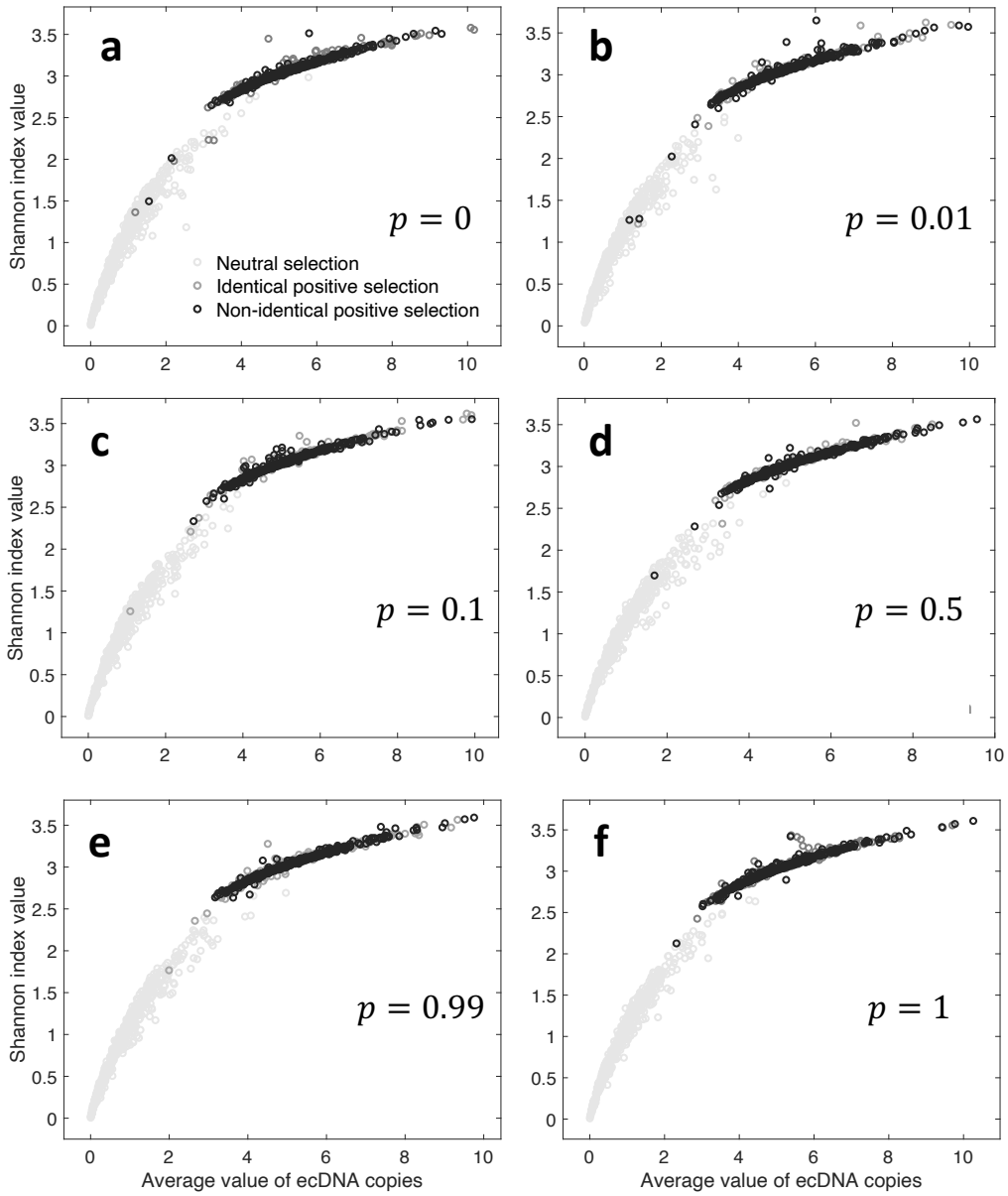


Figure S1: **Shannon index at whole population level for different selection scenarios and different values of phenotype switching ($p_y = p_r = p$).** Three selection cases are considered: neutral ($s_y = s_r = s = 1$), identical positive selection ($s_y = s_r = s = 2$) and non-identical positive selection ($s_y = 2$, $s_r = 1$). Each dot represents the outcome of a single simulation (500 identical repeated realisations per selection scenario in each panel).

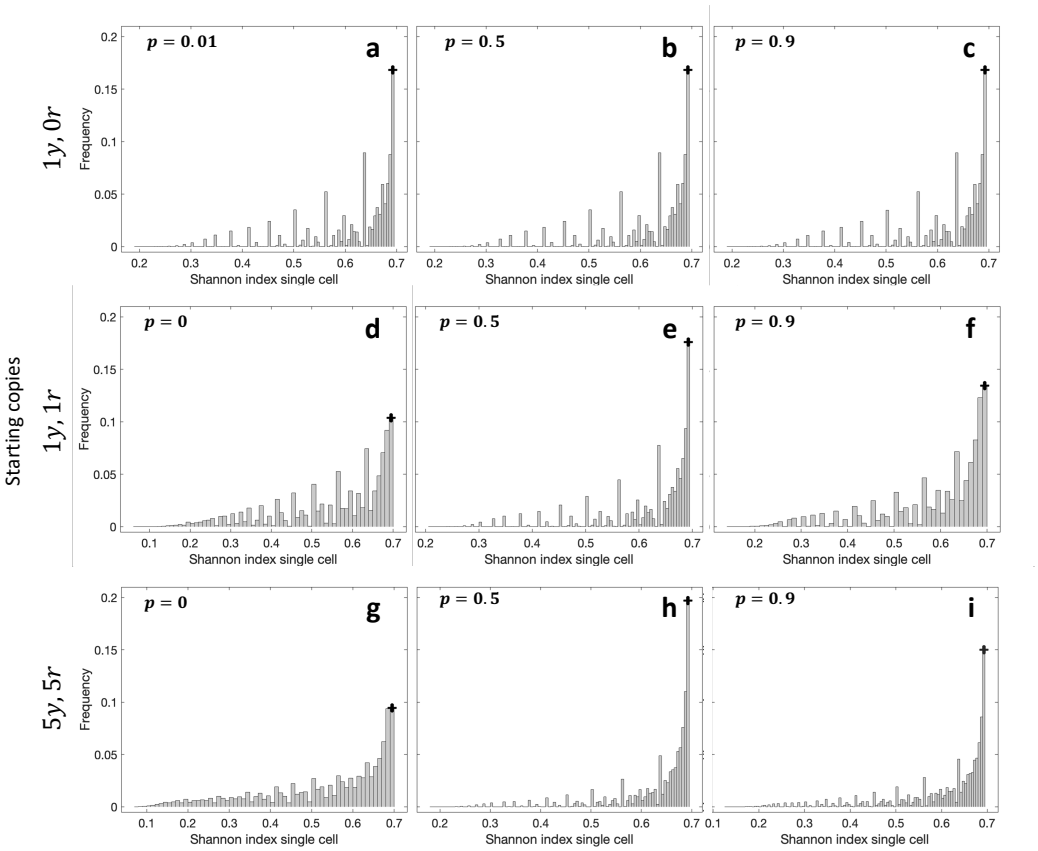


Figure S2: **Shannon index at single-cell level. Distribution of Shannon index values for different values of** p $(s_y = s_r = s = 2)$. The red dot on top of a column in each plot identifies the Shannon index value with maximum frequency, which is $H \simeq 0.635$ for every p. The initial configuration for these simulations is a single cell with 1 yellow and 1 red copy, and the indices are recorded at a population size of 10^4 cells over 600 realisations.

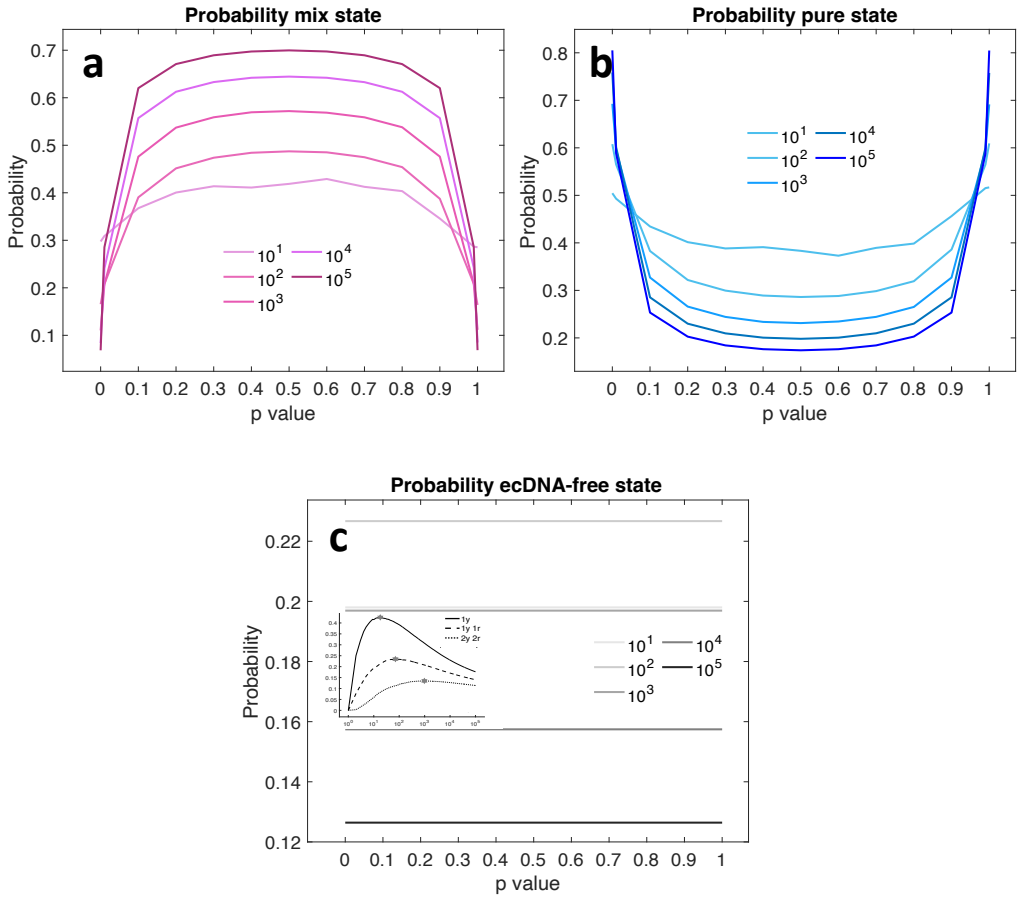


Figure S3: **Probability of staying in mix, pure and ecDNA-free states at different times (population size) under positive identical selection ($p_y = p_r = p$, initial condition: initial condition: 1 single mix cell with 1 yellow and 1 red copies.)**. **a.** Probability of staying in mix state. **b.** Probability of staying in pure state **c.** Probability of staying in ecDNA-free state. Furthermore, in the internal panel c, the trend of ecDNA-free cells over time for different initial conditions is shown. The star dots indicate the maximum value for ecDNA-free cells density for each curve. These dynamics are generated by simulations starting with a single cell with 1 yellow and 1 red ecDNA copies and results are recorded over 1000 realisations.

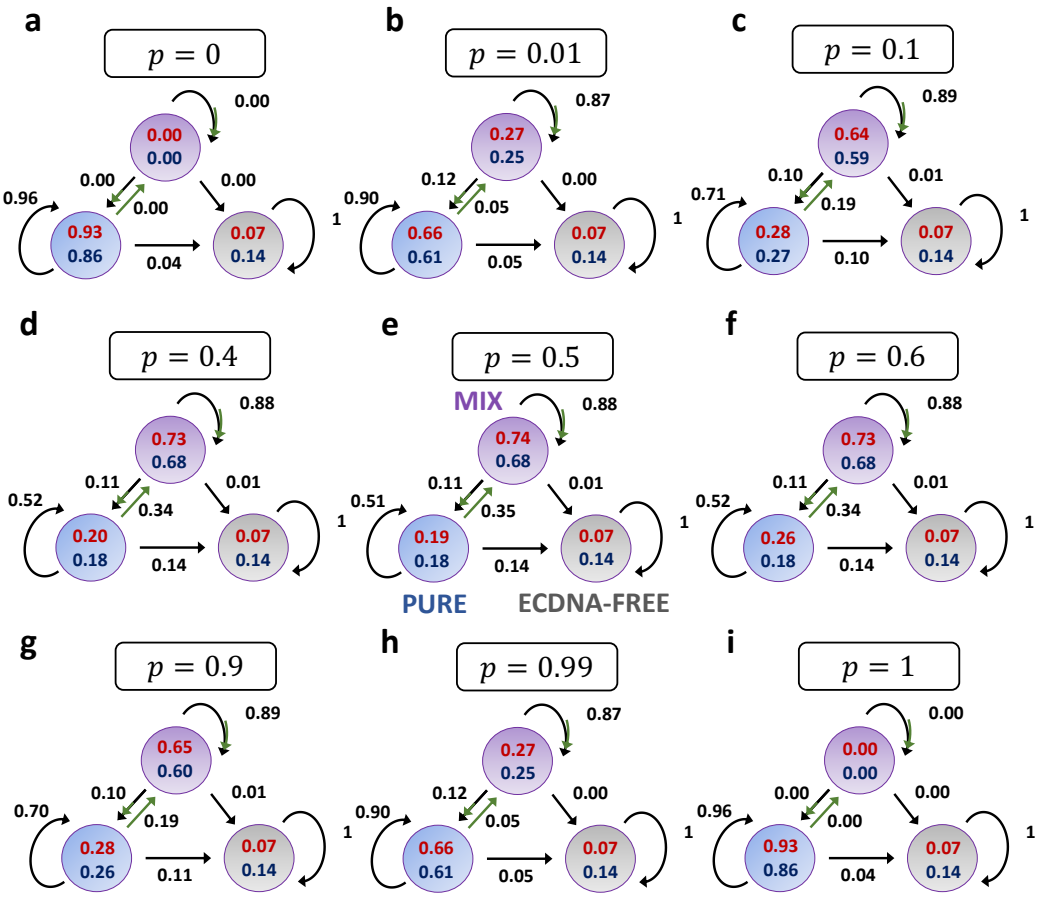


Figure S4: **Transition probabilities for model's dynamics under identical positive selection ($s_y = s_r = s$) recorded at a size of 10^5 cells ($p_y = p_r = p$, initial condition: 1 single pure cell with 1 yellow copy).** The above panels show the probability of transitioning from one state to another, i.e. the arrows' probability, and the one of staying in a state, i.e. the compartments' probability, when identical positive selection for both yellow and red ecDNA is considered. Specifically, two selection values are considered, respectively $s = 2$ and $s = 3.5$. Black arrows illustrate the transitions due to random segregation, whilst the green arrows the ones due to switching. The transition probabilities from one state to another, i.e. the arrows, do not change significantly for different positive selection values, whilst the probabilities of staying in a subpopulation, i.e. the compartments, are shown in red colour and blue colour respectively for $s = 3.5$ and $s = 2$. Regardless of the value of selection, when intermediate switching is considered (**d-f**), mix ecDNA state is the most favoured, with a higher probability of being in its state or transitioning from others, if compared to low (**b-c**) and high (**g-h**) values for p. When p is at its extreme values, meaning either $p = 0$ (**a**) or $p = 1$ (**i**), then we observe an identical behaviour of the dynamics, with the absence of cells in the mix ecDNA state and an high transition towards pure state. The initial configuration for these simulations is a single cell with 1 yellow ecDNA copy and results are recorded over 1000 realisations.

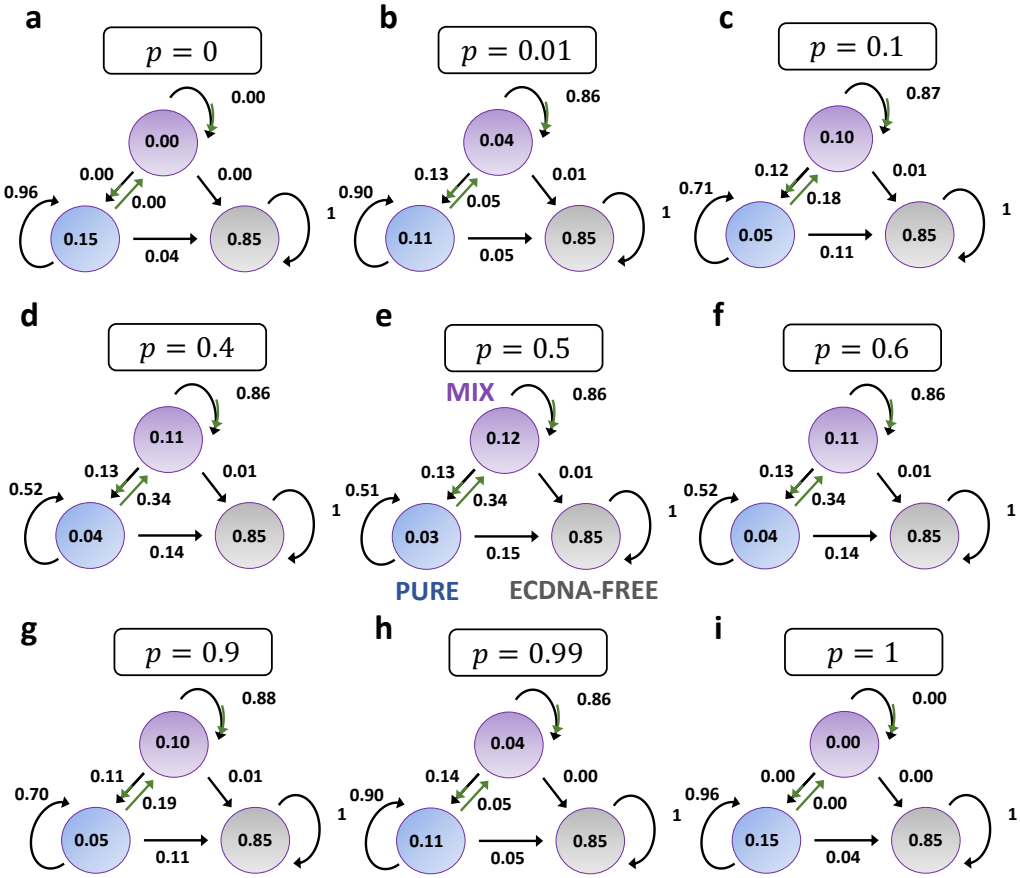


Figure S5: **Transition probabilities for model's dynamics under neutral selection, $s_y = s_r = s = 1$, recorded at a size of 10^5 cells ($p_y = p_r = p$, initial condition: 1 single pure cell with 1 yellow copy).** The above panels show the transition probabilities as in Figure S4 when neutral selection is considered. As in this selection scenario, ecDNA-free cells are the absorbing state, then the transitions and states probabilities change accordingly, compared to positive selection cases. The initial configuration for these simulations is a single cell with 1 yellow ecDNA copy and results are recorded over 1000 realisations.

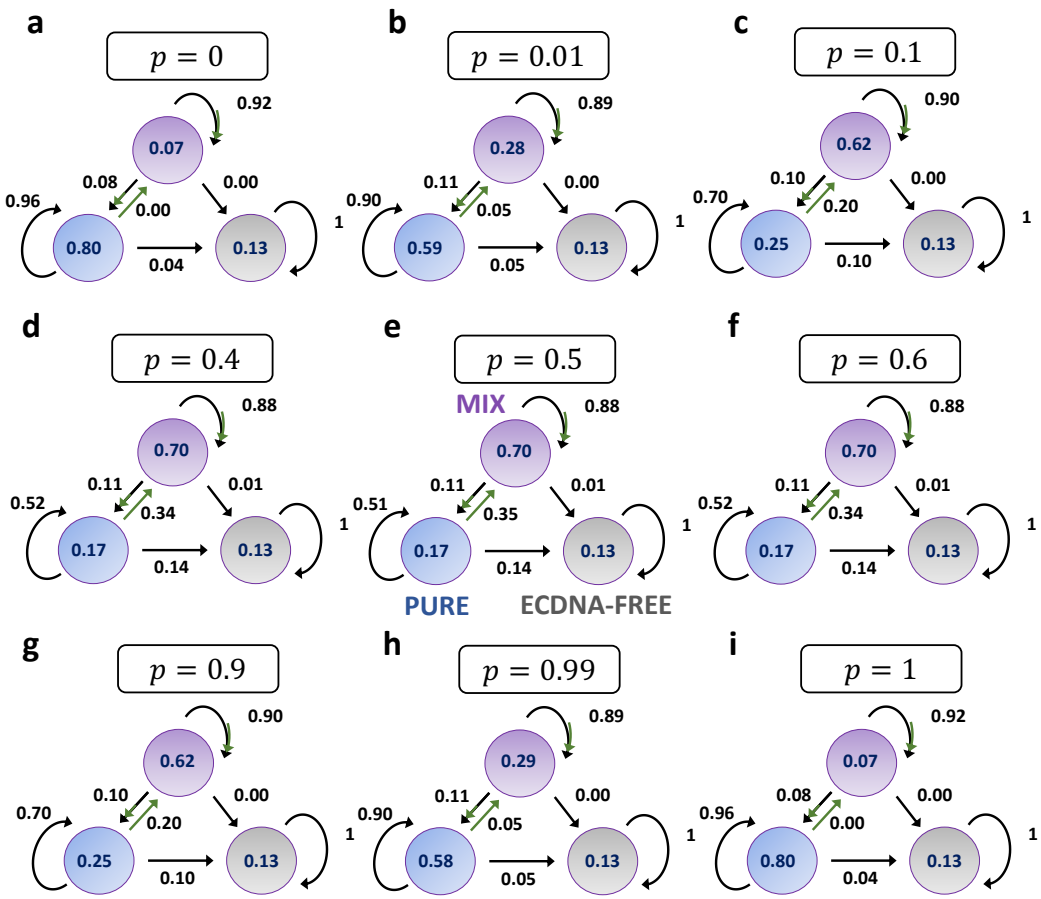


Figure S6: **Transition probabilities for model's dynamics under identical positive selection ($s_y = s_r = s = 2$) recorded at a size of 10^5 cells ($p_y = p_r = p$, initial condition: 1 single mix cell with 1 yellow and 1 red copies).** The above panels show the transition probabilities as in Figure S4 when another initial condition is considered. We can appreciate the same behaviour in terms of dynamics of mix cells, as they are more abundant under intermediate switching values (**d-f**). Also, the two extreme cases, i.e. $p = 0$ (**a**) and $p = 1$ (**i**), are symmetric. The initial configuration for these simulations is a single cell with 1 yellow and 1 red ecDNA copies and results are recorded over 1000 realisations.