Punctuational evolution is pervasive in distal site metastatic colonization

PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Butler G, Amend SR, Venditti C, Pienta KJ. 2025 Punctuational evolution is pervasive in distal site metastatic colonization. *Proc. R. Soc. B* **292**: 20242850.
https://doi.org/10.1098/rspb.2024.2850

Received: 20 September 2024
Accepted: 31 December 2024

Subject Category:
Evolution

Subject Area:
evolution

Keywords:
cancer evolution, metastasis, punctuational evolution

Authors for correspondence:
George Butler
e-mail: gbutle16@jh.edu
Kenneth J. Pienta
e-mail: kpienta1@jhmi.edu

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.7628654.

THE ROYAL SOCIETY PUBLISHING

Punctuational evolution is pervasive in distal site metastatic colonization

George Butler[1], Sarah R. Amend[1], Chris Venditti[2] and Kenneth J. Pienta[1]

[1]Cancer Ecology Center, The Brady Urological Institute, Johns Hopkins School of Medicine, Baltimore, MD 21287, USA
[2]School of Biological Sciences, University of Reading, Reading RG6 6AS, UK

GB, 0000-0002-6207-6225; CV, 0000-0002-6776-2355

The evolution of metastasis, the spread of cancer to distal sites within the body, represents a lethal stage of cancer progression. Yet, the evolutionary dynamics that shape the emergence of metastatic disease remain unresolved. Here, using single-cell lineage tracing data in combination with phylogenetic statistical methods, we show that the evolutionary trajectory of metastatic disease is littered with bursts of rapid molecular change as new cellular subpopulations appear, a pattern known as punctuational evolution. Next, by measuring punctuational evolution across the metastatic cascade, we show that punctuational effects are concentrated within the formation of secondary tumours at distal metastatic sites, suggesting that qualitatively different modes of evolution may drive primary and metastatic tumour progression. Taken as a whole, our findings provide empirical evidence for distinct patterns of molecular evolution at early and late stages of metastatic disease and our approach provides a framework to study the evolution of metastasis at a more nuanced level than has been previously possible.

1. Introduction

Cancer is an evolutionary process [1] in which genetic and epigenetic alterations accumulate over time, driving molecular divergence and widespread genetic intratumour heterogeneity [2]. As a result, multiple studies have sought to characterize the variation in the rate of evolution (the tempo) and the distribution of evolutionary rates through time (the mode) during primary tumour progression [3–7]. Yet, the primary cause of cancer-related mortality is owing to metastatic evolutions [8]—the spread of cancer to distal sites within the body. Despite its clinical importance, the evolutionary roadmap of metastatic disease remains poorly understood [9].

The spatial and temporal heterogeneity of metastatic disease creates both analytical and practical challenges for evolutionary characterization [10]. However, recent efforts with single-cell CRISPR-Cas9 lineage tracing technologies have helped overcome these issues, allowing for metastatic progression to be investigated at an unprecedented resolution and scale [11,12]. Yet, no study to date has measured the temporal diversity–the mode–in the evolution of metastatic disease at the single-cell level. For example, it is not clear whether the change in anatomical location from the primary tumour to a metastatic site is associated with increased heterogeneity in the distribution of molecular evolutionary rates through time. That is, given that only a subset of the initial cancer cell population will arrive at a distal site and change in the microenvironment, we might expect lineage diversification to be associated with molecular change, a dynamic known as punctuational evolution [13]. If true, then this might explain why metastatic tumours often have a higher degree of mutational clonality [14] and thus a lower degree of intratumour heterogeneity [9]. That is, a new, well-adapted clone may

arrive at a distal metastatic site and begin to proliferate, driving a burst of evolutionary change.

Punctuational evolution was originally proposed to explain the discontinuity of phenotypes within the fossil record [15]. Specifically, Eldredge and Gould coined the term *punctuated equilibria* in which evolutionary change is concentrated at the point of speciation followed by a period of evolutionary stasis [15,16]. In turn, subsequent experimental studies in bacteria showed that punctuational patterns of evolution at the phenotypic level can also be mirrored at the molecular level [17,18]. However, in contrast to punctuational evolution at the phenotypic level, punctuational evolution at the molecular level does not require a period of stasis, just an association between molecular divergence and lineage diversification [13]. Multiple studies have since leveraged branch-scaled phylogenetic trees to show that punctuational effects at the molecular level are widespread across species [13] and viral evolution [19]. Yet in cancer, punctuated-like patterns of evolution are typically considered as early events in tumour progression and are studied in the context of large-scale karyotypic alterations or copy number variations (CNVs) [20,21], thus potentially missing the presence of punctuational evolution driven by changes in the microenvironment as well as population size or structure. As a result, the presence of punctuational evolution, defined as the association between lineage divergence (*LD*) and lineage diversification in cancer, and specifically metastasis, remains unknown.

Here, we use single-cell lineage tracing data [12] in combination with a Bayesian phylogenetic framework [13] to test for evidence of punctuational evolution during metastatic progression. If punctuational effects are present, we expect to find a positive association between the amount of molecular divergence through time and the net increase in *LD*. In contrast, if punctuational effects are not present, we expect to find no association between the amount of molecular divergence through time and the net increase in *LD* (figure 1).

2. Results

(a) Reconstructing branch-scaled lineage trees at the single-cell level

All single-cell data used throughout are available from Yang *et al.* [12]. Briefly, sequencing data were collected from a lung adenocarcinoma genetically engineered mouse model (GEMM) in which embryonic stem cells were engineered with an evolving CRISPR-Cas9 dynamic lineage tracing system and inducible *Kras* and *Trp53* mutations, henceforth referred to as KP. Further data were also collected from two GEMMs that harboured an additional inducible LKB1 or APC mutation, henceforth referred to as KPL or KPA. Testing for punctuational effects across the three different mouse models controls for the potential effect of the individual GEMM system. The final dataset consisted of 10 mice, including 1 KP mouse, 7 KPL mice and 2 KPA mice. Across the 10 mice, target site information was recovered for 26 separate lineages (11 metastatic and 15 non-metastatic), spanning 71 individual tumours and 17 062 cells. The term 'lineage' is used throughout to refer to individual cells that are descendant from the same original embryonic stem cell (as verified from the Lenti-Cre-BC library [12]) (see §4).

To reconstruct the phylogenetic relationship in each lineage, we used the edited target site information in combination with a mixture-model-based approach of phylogenetic reconstruction fitted in a reversible jump Markov chain Monte Carlo (RJMCMC) framework [22]. The branch lengths in the reconstructed trees correspond to the amount of expected evolutionary divergence between pairs of individual diversification events (DEs), measured in units of nucleotide substitutions (figure 2*a*). The mixture-model-based approach is a modification of existing likelihood-based methods of CRISPR lineage reconstruction [23,24] as it allows for qualitatively different models of sequence evolution [22] to be estimated across different cut sites. This modification is important to control for the variability in the CRISPR editing process and to ensure that accurate branch lengths are recovered. In turn, we fitted a non-reversible substitution model [25] to account for the 'locked-in' nature of the CRISPR substitutions at a given target site [26]. Finally, we derived a sample of 1000 trees from the posterior distribution of each lineage to account for the inherent topological uncertainty in the estimation process [27] (see §4).

(b) Testing for evidence of punctuational evolution

To test for evidence of punctuational evolution, we tested for an association between the total *LD* and the total number of DEs (*DE*; one lineage splitting into two) [28]. The total *LD* was calculated by summing the individual branch lengths from root-to-tip for each cell. Likewise, the number of DEs was calculated by counting the number of nodes along the path from root-to-tip.

Next, we fitted a phylogenetic generalized least squares (PGLS) model in which *LD* was dependent on the number of DEs (*DE*) as well as the sum of the gradual effects in each branch along the path. That is, $LD = \beta * DE + g$, where β is the punctuational contribution to molecular divergence at each node and g is the gradual contribution. Under a gradual, homogenous model of evolution, we expect to find no association between *LD* and *DE* and thus $\beta = 0$. In contrast, under a punctuational model of evolution in which diversification events are associated with molecular change, we expect to find a positive association between *LD* and he number of diversification events *DE* and thus $\beta > 0$ (figure 1) [13].

We found a significant relationship between *LD* and *DE*s, and thus evidence of punctuational evolution, in 18 of the 26 lineages (see §4). However, 8 of the 18 lineages with evidence of punctuational evolution also suffered from a well-known artefact of phylogenetic reconstruction, known as the node density artefact (see §4d(ii); electronic supplementary material, figure S1) [29], and these were thus excluded from our analysis. As a result, our final dataset consisted of 18 lineages, 10 of which (7 metastatic and 3 non-metastatic) showed evidence of punctuational evolution (electronic supplementary material, table S1). Finally, in the seven metastatic lineages, we tested an alternative model in which a separate intercept and slope were estimated for metastatic cells within a given lineage. We found no significant difference between the estimated slopes and thus a single slope was estimated throughout (electronic supplementary material, table S2).

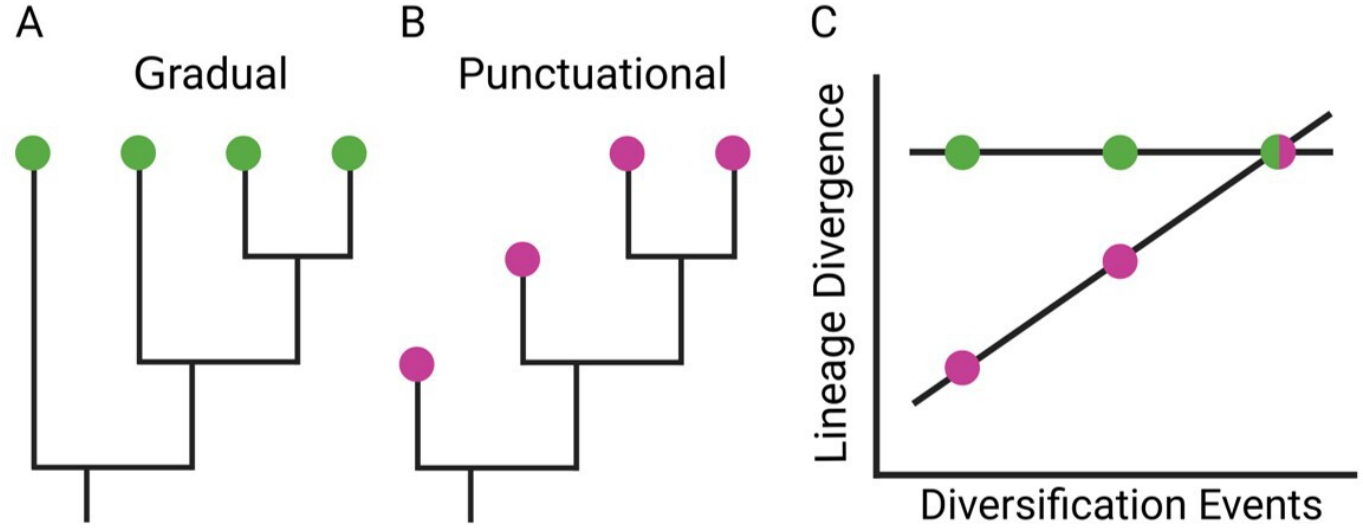

Figure 1. Detecting gradual and punctuational evolution. (A) The nodes in each tree represent a net increase in the number of cellular subpopulations. The branches (shown as vertical lines) represent the evolutionary change within a cellular subpopulation. The branching events (where one branch splits into two) represent the splitting of a subpopulation into two. A gradual model of evolution assumes that molecular change, represented by branch lengths on the phylogenetic tree, is independent of lineage diversification. (B) A punctuational model of evolution assumes that molecular change occurs at the point of lineage diversification. (C) The relationship between lineage divergence—the sum of the branch lengths from root-to-tip—and the number of diversification events along the path can be used to test for evidence of punctuational evolution. A gradual model of evolution predicts no association between lineage divergence and diversification events. In contrast, a punctuational model predicts a positive association between lineage divergence and diversification events.

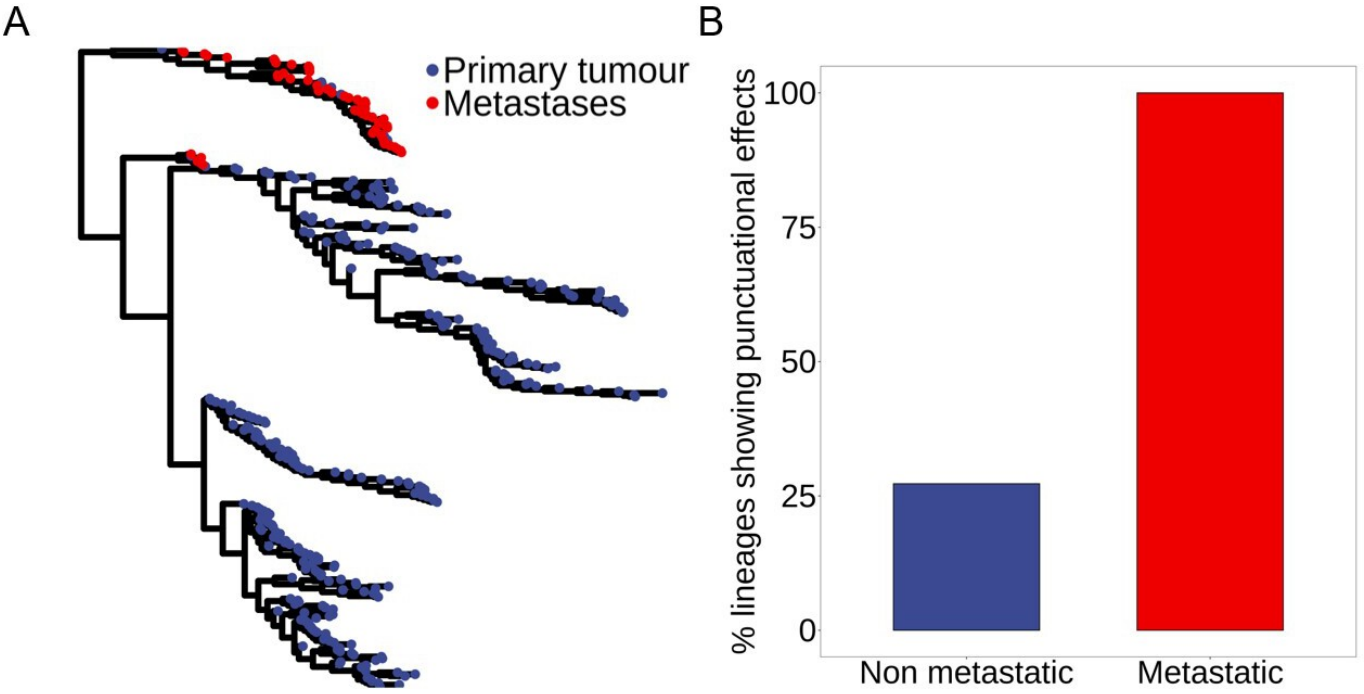

Figure 2. Punctuational evolution is more frequent in metastatic lineages. (A) A metastatic single-cell lineage tree. The blue and red points correspond to the individual cells recovered from the primary tumour and distal site metastases, respectively. The branch lengths are proportional to the estimated kappa-transformed ($\kappa = 0.32$) expected evolutionary divergence measured in units of nucleotide substitutions. (B) The dataset consists of 18 lineages (7 metastatic and 11 non-metastatic). Punctuational evolution is significantly more common in metastatic compared to non-metastatic lineages ($p = 0.011$). All seven metastatic lineages show evidence of punctuational evolution but only three out of 11 non-metastatic lineages showed evidence of punctuational evolution (see electronic supplementary material, Table S1).

Evidence of punctuational evolution was present in both metastatic and non-metastatic lineages ranging in size from 247 to 1612 cells. However, we found that punctuational evolution was significantly more common in metastatic rather than non-metastatic lineages ($p = 0.011$, figure 2*b*). In fact, we found evidence of punctuational evolution in all 7 metastatic lineages compared with only three of the 11 non-metastatic lineages. In contrast, while evidence of punctuational evolution was present in all three GEMMs (KP, KPL and KPA), we found no association between the presence of punctuational evolution and the specific mouse model ($p = 0.207$, electronic supplementary material, figure S2), suggesting that punctuational effects are independent of the given experimental mouse model.

Given that punctuational-like patterns of evolution in cancer are commonly considered with respect to karyotypic alterations and CNVs [20,21], we fitted a secondary model in which the number of CNVs was included as an additional independent

variable (see §4). We found no evidence of a significant association between LD and CNVs, yet the number of DEs remained significant in the ten lineages that had been identified previously (electronic supplementary material, table S3). Taken as a whole, these results suggest that punctuational effects are associated with metastatic progression and that punctuational patterns of evolution are not strictly a copy number-based phenomenon in cancer progression [30]. We postulate that the apparently discordant role of CNVs in punctuational evolution may be a consequence of the phylogenetic relationships in previous studies being inferred from the copy number profile itself. As a result, unravelling the effect of copy number alterations compared to the effect of diversification more broadly may not have been possible.

The slope between LD and the number of DEs (β) captures the increase in branch length (amount of molecular evolution) per DE [13]. Yet, the absolute value of β will vary depending on the lineage-specific overall rate of molecular evolution. That is, certain lineages may have a higher overall rate of molecular evolution compared with others. To allow for cross-lineage comparisons, we used β to estimate the proportion of the total molecular diversity within each lineage that is attributed to punctuational effects (see §4). Across the 10 lineages with evidence of punctuational evolution, the proportion of molecular diversity attributed to punctuational effects ranged from 0.011 to 0.125, with an average value of 0.06 (electronic supplementary material, table S1). That is, when punctuational evolution is present, an average of 6% of the molecular diversity is attributed to punctuational effects.

(c) Punctuational evolution has a greater impact on tumour progression at distal site metastases

Punctuational effects quantify the heterogeneity in evolutionary rates through time associated with lineage diversification. Two potential mechanisms of lineage diversification that may explain the presence of punctuational evolution are founder effects and ecological niche invasion [13].

Founder effects refer to instances in which a subset of the population becomes spatially separated causing a temporary amplification in the effect of genetic drift prior to population expansion [31]. In contrast, ecological niche invasion refers to adaptive evolution in which a subset of the population is better adapted to a set of new ecological stressors, leading to a temporary increase in the rate of evolution at the point of lineage diversification [32]. Crucially, both mechanisms are temporary [33] and both have the potential to act during cancer evolution, and specifically during the evolution of tumours at distal site metastases. For instance, the number of cells that arrive at a distal site (generally thought to be a single cell or small cluster of cells) is expected to be only a fraction of the total within the primary tumour (billions) [34]. Likewise, the environmental selective pressures at a distal site are expected to be different from the primary tumour microenvironment in which the cells first evolved [35,36]. As a result, we hypothesize that punctuational evolution would have a greater impact on the evolutionary trajectory of tumour formation at a distal site compared with inside the primary tumour. If true, we would expect to find a greater punctuational contribution to molecular diversity at a distal site compared with within the primary tumour.

To test whether the punctuational contribution to molecular diversity was different in the primary tumour compared with distant site metastases, we partitioned the branches in each metastatic lineage according to whether the evolutionary change had occurred in the primary tumour or in a distal metastatic site. That is, each cell within a given lineage was labelled by location: primary tumour or distal metastatic site (e.g. lymph node, liver and bone). In turn, for each branch in each tree in the posterior distribution for a given lineage, we compared the states of the ancestral and descendant nodes using a probabilistic method of ancestral state reconstruction (ASR) to account for the uncertainty in the estimation process [37]. If both the ancestral and descendant nodes were estimated to be in the same location, then the branch was categorized as that location (i.e. either primary tumour or metastatic). In contrast, if ancestral and descendant nodes were estimated to be in different locations, then the branch was categorized as a transition (figure 3*a*). Finally, we quantified the primary tumour- and metastatic-specific punctuational contribution within each lineage by calculating the ratio of primary tumour or metastatic branches with respect to the proportion of the total tree length attributed to either the primary tumour or distal site metastases (see §4).

We found that the punctuational contribution to molecular diversity was significantly higher in distal site metastases compared to the paired primary tumour in six of the seven metastatic lineages (figure 3*b*, electronic supplementary material, table S4). Moreover, we found that the proportion of molecular diversity attributed to punctuational effects ranged from 0.011 to 0.162 in distal site metastases compared to only 0.005–0.061 in the primary tumour. That is, 1–16% of the molecular diversity in distal site metastases is attributed to punctuational effects compared to only 0.5–6% of the diversity within the primary tumour. These results therefore suggest that punctuational evolution has a greater impact on the molecular landscape of distal site metastases compared with inside the primary tumour.

Owing to the limited number of lineages with metastatic tumours of substantial size at different anatomical sites, we were unable to test whether punctuational effect sizes were differentially associated with specific metastatic sites. However, given that previous multi-region studies have shown that the lymph-node metastases emerge through a wider evolutionary bottleneck compared with metastases at other distal sites [34], we would speculate that the punctuational contribution in lymph-node metastases would be smaller compared with other distal sites. Future investigation in this area is likely to result in further significant contributions to cancer biology. Nevertheless, these results here represent the first empirical evidence for qualitatively different modes of evolution at early and late stages of metastatic progression.

(d) Complex patterns of metastatic seeding are associated with stronger punctuational effects

Metastatic dissemination is a complex process in which only a fraction of the cells that leave the primary tumour ultimately seed and colonize at a distant site [38,39]. A growing body of evidence suggests that patterns of metastatic seeding are

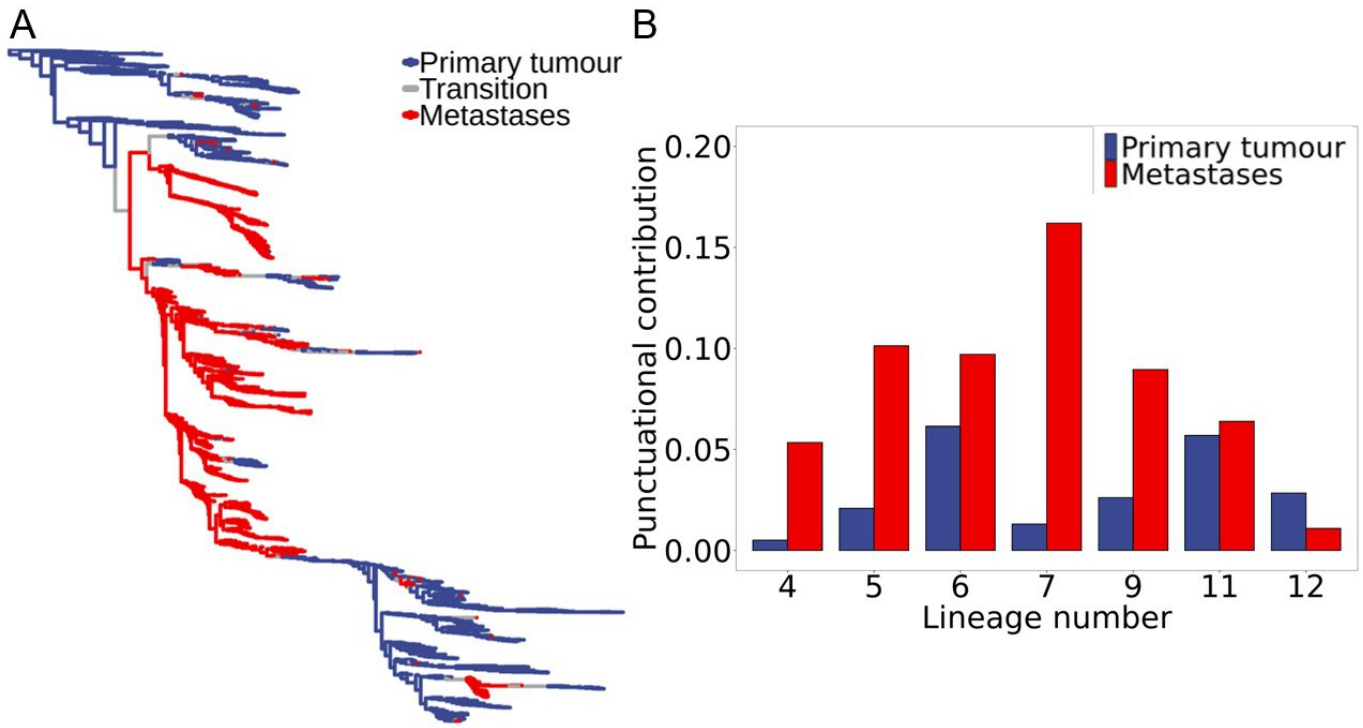

Figure 3. Punctuational effects contribute more to the molecular diversity in distal site metastases. (A) A metastatic single-cell lineage tree with inferred ancestral anatomical locations. The tree is transformed according to the estimated kappa value ($\kappa = 0.31$). The blue and red branches correspond to evolutionary change estimated to have occurred in the primary tumour and distal site metastases, respectively. The grey branches correspond to evolutionary change that has occurred in which the ancestral and descendent cells are estimated to be in different anatomical locations. (B) The punctuational contribution to molecular diversity is significantly higher in distal site metastases (shown in red) compared to the paired primary tumour (shown in blue) in six of the seven metastatic lineages (electronic supplementary material, table S4).

considerably more complex than previously appreciated [11,40,41]. Specifically, multiple studies have shown that a given metastatic site can be seeded by multiple independent lineages [42] and that cells from metastatic tumours can subsequently seed new metastases at different sites within the body [43]. Given that punctuational effects cause deviations from gradual evolution [13], and our results show that punctuational effects have a greater impact on the trajectory of metastatic tumours, we speculate that more complex patterns of metastatic dissemination would result in stronger punctuational effects and thus greater deviations from gradual evolution.

To test the hypothesis that metastatic lineages seeded by a single monophyletic branch will have a smaller deviation from gradual evolution compared with lineages that have polyphyletic seeding, we quantified the size of the deviation from gradual evolution by calculating the correlation coefficient, ρ, between LD and diversification (see §4). We then characterized the level of 'metastatic complexity' by estimating the degree of phylogenetic dispersion, D [44], in the location of cells recovered from the primary tumour and distal site metastases. High metastatic complexity refers to instances in which metastatic cells are distributed throughout the tree (figures 4*a* and 3*a*). In contrast, low metastatic complexity refers to instances in which metastatic cells are clustered within a similar area(s) of the tree (figures 4*b* and 2*a*).

First, we fitted a model across both metastatic and non-metastatic lineages in which deviations from gradual evolution are dependent on the size of the punctuational contribution. We found a significant positive association between the deviations from gradual evolution and punctuational contributions ($p < 0.001$, figure 4*c*). That is, stronger punctuational effects cause greater deviations from gradual evolution [13]. Yet, along the punctuational continuum, we also found that greater deviations from molecular gradualism were present in lineages with increased metastatic complexity (figure 4*c*), suggesting that more complex patterns of metastatic seeding and reseeding are associated with increased heterogeneity in rates of molecular evolution through time.

3. Conclusion

The evolution of metastasis signifies the evolution of lethal cancer. Yet despite its clinical importance, the evolutionary trajectory of metastatic disease remains broadly unknown [9]. Here, we combined single-cell lineage tracing data [12] with robust phylogenetic comparative methods [13] to reveal that punctuational effects are a common molecular feature of metastatic progression. Furthermore, by measuring punctuational effects throughout the disease course, we show that punctuational evolution has a greater impact on the evolutionary trajectory of distal site metastases compared to primary disease. Taken as a whole, these results represent the first empirical evidence of differential modes of evolution at early and late stages of metastatic progression and highlight the need to study metastatic dissemination as a continuous process rather than disjoint stages. Moreover, these results provide a tantalizing window into the breadth of evolutionary dynamics that develop as a function of

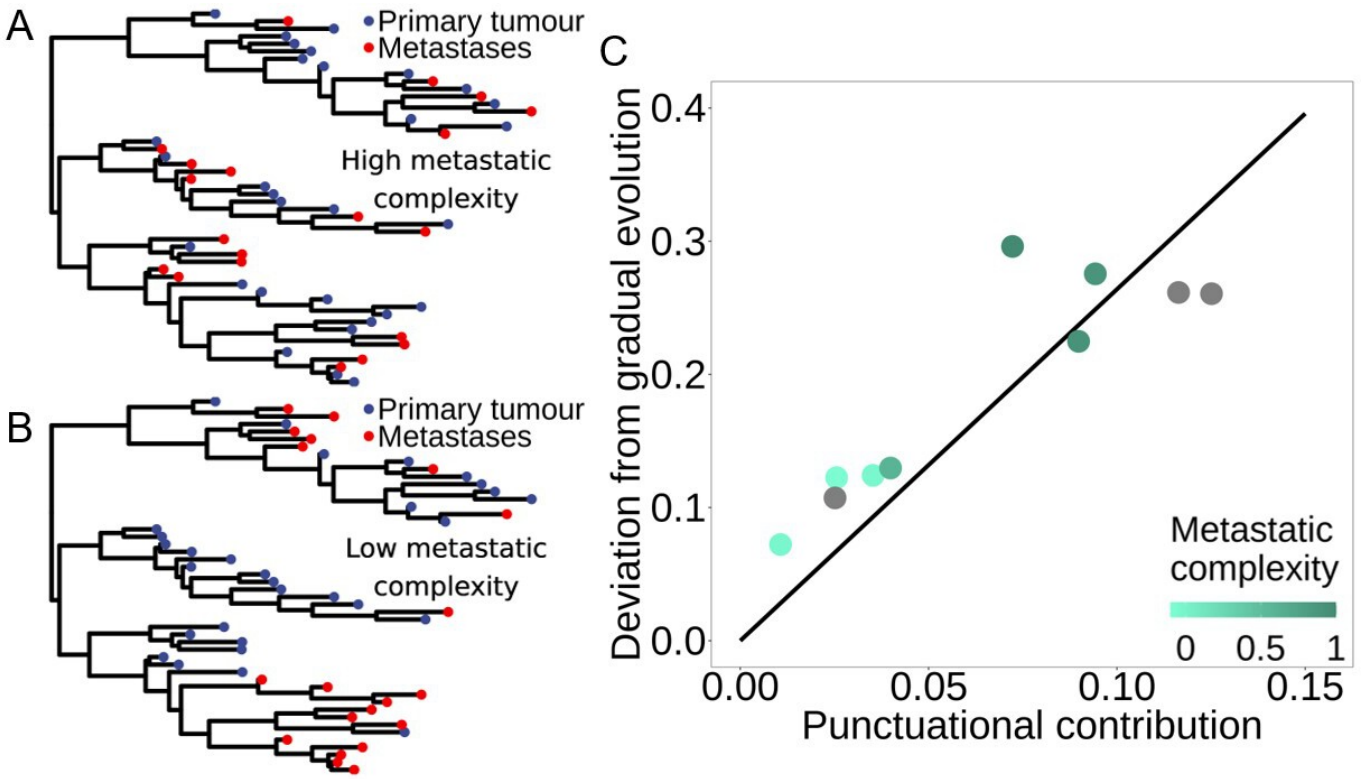

Figure 4. Larger deviations from gradual evolution are associated with stronger punctuational effects and increased metastatic complexity. (A) A simulated tree with high metastatic complexity ($D = 0.97$) in which both the red and blue tips are distributed across the tree. (B) A simulated tree with low metastatic complexity ($D = -0.01$) in which the red and blue tips are clustered within the tree. In both trees, the blue and red points correspond to simulated cells recovered from the primary tumour and distal site metastases. (C) Metastatic lineages are shown in shades of turquoise and non-metastatic lineages are shown in grey. Darker shades of turquoise are associated with increased metastatic complexity, as measured by the degree of phylogenetic dispersion, D, in the location of cells within a given lineage that were recovered from the primary tumour and distal site metastases (see §4). The fitted line shown in black is estimated from both metastatic and non-metastatic lineages and shows a significant positive association between the deviation from gradual evolution and the magnitude of the punctuational contribution ($p < 0.001$). Furthermore, lineages with increased metastatic complexity (shown in darker turquoise) are associated with larger deviations from gradual evolution.

the interplay between a cancer cell and its environment [36], paving the way for future studies to investigate how exogenous factors such as systemic therapy differentially shape the trajectory of lethal disease progression. Finally, single-cell barcoding techniques provide an increased level of cellular and temporal granularity beyond what was previously possible from existing bulk and endogenous single-cell lineage tracing studies. As a result, we hope that our findings and approach stimulate further development of analytic techniques that seek to leverage the wealth of information that is obtained from different barcoding technologies.

4. Methods

(a) Data and code availability

Raw single-cell Target Site (for lineage reconstruction) and RNA-seq (for measuring copy number status) libraries as well as processed single-cell clonality and sample information (verified from MULTI-seq and Lenti-Cre-BC libraries) are available from Yang *et al.* [12]. Briefly, sequencing data were collected from a lung adenocarcinoma GEMM in which embryonic stem cells were engineered with an evolving CRISPR-Cas9 dynamic lineage tracing system and inducible *Kras* and *Trp53* mutations, henceforth referred to as KP. Further data were collected from two GEMMs that harboured additional inducible LKB1 or APC mutation, henceforth called KPL and KPA. Information regarding library construction, embryonic stem cell engineering and sample preparation is detailed in Yang *et al.* [12].

All original code is available on GitHub at https://github.com/george-butler/punctuational_evolution_cancer. The software BayesPhylogenies [22] used for lineage reconstruction is available at https://www.evolution.reading.ac.uk/BayesPhy.html. The software BayesTraitsV4 [45] used to test for evidence of punctuational evolution and reconstruct ancestral states is available at https://www.evolution.reading.ac.uk/BayesTraitsV4.1.1/BayesTraitsV4.1.1.html.

(b) Single-cell data preprocessing

(i) Target site preprocessing

Target site libraries were processed using a modified version of the Cassiopeia [46] preprocessing pipeline (v. 2.0.0) to retain the 296 base pair target sites in the final output table. The Cassiopeia parameters and threshold values were kept the same as outlined in Yang *et al.* [12]. The preprocessed clonality and sample information (outlined above) were used to assign each of the target sites to a given lineage. The term 'lineage' is used throughout to refer to individual cells that are descendant from the same original embryonic stem cell (as verified from the Lenti-Cre-BC library). To be included in the final dataset, a given lineage needed to have been recovered from a mouse that harboured at least one metastatic and one non-metastatic lineage. This criterion was used to try to control for between-mouse differences that may affect the likelihood of metastatic progression. The final dataset consisted of 10 mice, including 1 KP mouse, 7 KPL mice and 2 KPA mice. Across the 10 mice, 26 separate lineages were recovered (11 metastatic and 15 non-metastatic), spanning 71 individual tumours and 17 062 cells.

(ii) Target site alignment for phylogenetic reconstruction

Prior to phylogenetic reconstruction, the target site information for each lineage was translated into a NEXUS file format. First, the 14-base-pair random integration barcode was removed from each target site, as well as the preceding 20 base pairs, resulting in a 262-base-pair sequence containing the 3 individual cuts. On average, 10 unique target sites were recovered for each lineage (19 out of 26 lineages had 10 unique target sites) and all lineages had at least 7 unique target sites. However, owing to practical limitations, the number of target sites that were recovered for an individual cell within a given lineage varied and may have been lower than the maximum number of target sites within the lineage. For instance, 10 unique targets were recovered for a given lineage but only 7 of the 10 target sites were recovered for a specific cell within the lineage. As a result, the missing target sites for an individual cell were marked with a blank sequence to preserve indexing across the alignments. Finally, the target sites for each cell within a given lineage were concatenated into a single N*262 sequence, where N is equal to the number of unique target sites in the lineage, and the integration barcodes were used to preserve the ordering across the lineage.

(iii) Measuring copy number variations

CNVs were quantified as outlined in Yang *et al.* [12]. Briefly, RNA-seq libraries were processed using the 10× CellRanger pipeline (v. 2.1.1) with the mm10 genome build. CNVs were then inferred from the processed RNA-seq libraries via the InferCNV R package (v. 1.10.1) in which the parameters and threshold values were kept the same as outlined in Yang *et al.* [12]. All tumours containing a minimum of five cells were processed independently, with normal lung cells from each background (KP, KPL and KPA) used as a reference. The number of CNVs per cell was calculated by summing the number of predicted CNV regions. In total, the CNV status was quantified for 17 056 out of 17 062 cells for which target site information was present.

(c) Lineage reconstruction

Lineage reconstruction is an analytically challenging process owing to the complex evolutionary dynamics, such as homoplasy, that can emerge over time within a population. However, CRISPR-Cas9 lineage tracing data also have a unique set of characteristics that further complicate the reconstruction process. For instance, once edited, substitutions are considered to be 'locked in' at a given cut site. Second, substitution rates can vary greatly between different target sites, as well as between individual cut sites within a given target site. Finally, large deletions can cause adjacent cut sites to be overwritten, thus erasing previous substitutions [26,47,48]. Taken together, these characteristics create a heterogeneous set of dynamics that need to be properly accounted for during the reconstruction process to ensure that accurate topologies and branch length estimates are recovered.

(i) Capturing complex patterns of molecular evolution in lineage tracing data

To detect and account for the heterogeneous patterns of evolution in CRISPR-Cas9 lineage tracing data we used a mixture-model based approach within a RJMCMC framework [22].

First, given that the starting unedited target site sequence is known, we fitted a general time non-reversible (GTNR) substitution model to estimate the frequency and rate of substitutions at sites within the target site. The GTNR model has an asymmetric 4×4 substitution rate matrix, $\mathbf{Q}$, allowing for independent rates of substitution to be estimated in each direction. That is, a different substitution rate can be estimated from A to C compared with C to A [25,49]. This detail is important as it allows for maximal flexibility and accounts for the 'locked in' nature of substitutions at a given target site [26]. Next, to account for faster and slower rates of substitution at individual sites, we used a four-part discrete gamma rate model [50] to scale the individual elements of the substitution rate matrix, $\mathbf{Q}$.

To account for qualitatively different patterns of molecular evolution across different areas of the alignment, such as between different target sites or cut sites, we used a mixture-model approach to fit j different evolution evolutionary models. That is, we estimated j-independent $\mathbf{Q}$ matrices and then calculated the weighted sum to find the likelihood at each position across the sequence alignment. Thus, in theory, an individual model of evolution can be estimated for each cut site, or even different

parts of each cut site in the alignment. Finally, to determine how many **Q** matrices need to be estimated within a given lineage, we used a RJMCMC framework. The RJMCMC framework removes the need for *a priori* specification and instead simultaneously estimates the elements of each **Q** matrix while also estimating the number of individual **Q** matrices. In turn, this reduces model complexity and removes the need for *post-hoc* model comparisons, thus ensuring that reconstruction is tractable for lineage(s) spanning thousands of individual cells. Further details regarding the mixture-model-based approach and the RJMCMC framework are detailed in [22].

(ii) Bayesian lineage reconstruction

Lineage reconstruction was performed using the software BayesPhylogenies [22]. The original unedited target site sequence was set as an outgroup to root each of the 26 lineages. Default priors were used throughout and the MCMC chains were sampled every 100 000th iteration after visual convergence was achieved. A 1000 tree posterior distribution was produced for each lineage and the outgroup was removed prior to all subsequent stages of analysis. We found that the number of **Q** matrices estimated for each lineage scaled with lineage size and ranged between 1 and 24, highlighting the heterogenous mix of evolutionary patterns that emerge over time and underlining the importance of estimating different models of evolution.

(d) Detecting punctuational evolution

To test for evidence of punctuational evolution, we tested for an association between total lineage divergence, *LD*, and the number of net diversification events, *DE*. Specifically, we used the equation $LD = \beta \times DE + g$, where β is the punctuational contribution of lineage diversification to molecular evolution at each node and g is the gradual contribution, as described in Pagel *et al.* [13]. Under a gradual model of evolution, we expect to find no association between *LD* and *DE*s. Thus, $\beta = 0$ and the equation simplifies to $LD = g$, meaning that a gradual model of evolution is sufficient to explain the lineage divergence over time. In contrast, under a punctuational model of evolution, we expect to find a positive association between lineage divergence and diversification, and thus $\beta > 0$ *(13)*.

(i) Quantifying punctuational effects

LD was calculated for each cell by summing the individual branch lengths from root-to-tip for each tree in the posterior distribution using the distRoot function in the adephylo R package [51]. Similarly, the distRoot function was also used to count the number of *DE*s along the path from root-to-tip for each cell in each tree in the posterior distribution. A PGLS model was fitted in a maximum likelihood framework using the software BayesTraits [45] to estimate β separately for each tree in the posterior distribution. Phylogenetic signal (λ) was estimated during the model-fitting process [52]. Default parameters were used throughout and each tree was scaled to have a mean branch length equal to 0.1.

(ii) Node density artefact

The node density artefact is a systematic error of phylogenetic reconstruction that can cause branch lengths to be underestimated in areas of the tree with fewer taxa [29]. As a result, the node-density artefact can bias regression estimates and cause a spurious relationship to appear between the amount of inferred total lineage divergence, *LD* and the number of net diversification events, *DE*.

To check for the presence of the node density artefact, we used the δ test that predicts a curvilinear relationship between *LD* and *DE* when the artefact is present (electronic supplementary material, figure S1) [28]. Specifically, the δ test fits a model of the form $LD = \beta \times DE^{1/\delta} + g$, where *LD* is the lineage divergence, β is the punctuational contribution, *DE* is the number of DEs, δ is the exponential DE scaler and g is the gradual contribution. We used the software BayesTraits [45] to estimate β and δ simultaneously for each tree in the posterior distribution. Default parameters were used throughout, and each tree was also scaled to have a mean branch length equal to 0.1. The node density artefact is considered present when β is significantly bigger than 0 and δ is numerically greater than one [13,53].

(iii) Criteria for identifying lineages with punctuational effects

We classified a lineage as having evidence of the node density artefact if $\beta > 0$ and $\delta > 1$ (see §4d(ii)) in at least 50% of the posterior distribution [13]. A total of seven lineages (three metastatic and four non-metastatic) were found to have evidence of the node-density artefact and were thus removed from subsequent stages of analysis. In the remaining 18 lineages, we classified a lineage as having evidence of punctuational evolution if $\beta > 0$, $\delta \leq 1$ (see §4d(ii)) in at least 50% of the posterior distribution and that the average value of δ across the posterior distribution was <1. We found evidence of punctuational evolution in 10 out of the 18 lineages (7 metastatic and 3 non-metastatic).

(iv) Assessing regression parameter significance

Regression parameter significance was assessed by the proportion of trees in the posterior distribution in which the parameter was significant at the 5% level. A parameter was classified as significant if in at least 50% of the trees in the posterior distribution the parameter was significant at the 5% level [13].

(e) Quantifying the punctuational contribution to molecular diversity

The punctuational contribution to molecular diversity was quantified as outlined in Pagel *et al.* [13]. Briefly, a bifurcating tree has $2(c-1)$ branches, where c is equal to the number of individual cells in the lineage. If the overall length of the tree, T, is equal to the sum of the individual branch lengths measured in unit nucleotide substitutions, then the ratio $2(c-1)\beta/T$ measures the proportion of the tree length,–the total molecular diversity–attributed to punctuational effects where β is the punctuational contribution of lineage diversification to molecular evolution at each node. Thus, if all LD is owing to gradual effects, then $\beta = 0$ and the punctuational contribution, $2(c-1)\beta/T$, also equals 0. In contrast, if no gradual effects are present, $2(c-1)\beta = T$ and the ratio $2(c-1)\beta/T = 1$.

(f) Quantifying the punctuational contribution across the metastatic cascade

To test whether the percent punctuational contribution to molecular diversity was different in the primary tumour compared to within distant site metastases, we partitioned the branches in each metastatic lineage according to whether the evolutionary change had occurred in the primary tumour or a distal metastatic site. That is, for each branch in each tree in the posterior distribution for a given lineage we compared the state of the ancestral and descendant nodes using ASR (see §4f(i)). Once finished, the branch lengths were then grouped into one of four categories: primary tumour, distant site metastases, transition, or undefined (see §4f(i)). Finally, we quantified the primary tumour- and metastatic-specific punctuational contribution within each lineage and then tested for a significant difference by comparing the pairwise difference between the two posterior distributions (see §4f(iii)).

(i) Inferring ancestral states

To infer the state of each internal node within the posterior distribution of a given lineage we fitted a continuous time Markov model [37] using the location from which each cell in the lineage had been recovered, e.g. primary tumour or distal metastatic site (any tumour that is not the primary tumour, e.g. lymph node, liver and bone). The model was fitted in a maximum likelihood framework using the software BayesTraits [45] to simultaneously estimate the transition rate from the primary tumour (P) to a distant metastatic site (M) (q_{PM}), and the reverse (q_{MP}), while also reconstructing the ancestral states for each node in the tree. We fixed the state of the root equal to the primary tumour state (P) but set no *a priori* restrictions on either of the transition rates (q_{PM} and q_{MP}). The phylogenetic parameter kappa (κ) was estimated for each tree during the model fitting process to transform the individual branch lengths [52]. Default parameters were used throughout, and each tree was scaled to have a mean branch length equal to 0.1.

(ii) Categorizing anatomical branch locations

We categorized branches by comparing the ancestral and descendant nodal states resulting from the ASR procedure. The state of an internal node was set equal to the state with the highest probability. If both states had equal probability, $P_{ASR} = 0.5$, then the nodal state was marked as 'undefined'. If both the ancestral and descendant nodal states were the same (e.g. ancestor = P and descendant = P or ancestor = M and descendant = M) then the branch was also assigned to the same state. If the ancestral and descendant nodal states were different (e.g. ancestor = P and descendant = M or ancestor = M and descendant = P) then the branch was assigned as a transition. Finally, if either the ancestral or descendant nodal state was 'undefined' then the branch was assigned as undefined.

(iii) Comparing punctuational contributions within a lineage

We quantified the primary tumour and metastatic punctuational contribution within a given lineage by calculating the ratio of primary tumour or metastatic branches with respect to the proportion of the total tree length attributed to either the primary tumour or distal site metastases. The formula to calculate the punctuational contribution is outlined above (see §4e). To be included in the subsequent comparison, each tree needed to have at least 10 primary tumour branches and 10 metastatic branches, and the estimated punctuational contribution needed to be less than 1. The terminal branches were excluded owing to their short length and thus their potential to bias the average punctuational contribution. However, the same qualitative results are obtained when the terminal branches are included. Finally, we used a multiple comparison test with a Bonferroni correction to compare the punctuational contribution between the primary tumour and distal site metastases.

(g) Quantifying deviations from molecular gradualism

We quantified the size of the punctuational departure from gradual evolution by estimating the correlation, ρ, between LD and the number of DEs as outlined in Pagel *et al.* [13]. Specifically, $\rho = (\beta\sigma^2_{LD} + \sigma_{g,LD}) / ((\beta^2\sigma^2_{LD} + \sigma^2_g)^{½} (\sigma^2_{LD})^{½})$ where σ^2_{LD} is equal to the variance in the number of lineage DEs, σ^2_g is equal to the variance in the gradual effects and $\sigma_{g,LD}$ is equal to the covariance between the gradual component and the number of lineage DEs, and is assumed to be zero. If no punctuational effects are present, $\beta = 0$, we do not expect to find any deviations from a clock-like tempo of evolution and thus no correlation between LD and diversification. In contrast, if all molecular divergence is wing to the punctuational effects, then we expect to find a perfect positive correlation between LD and diversification. That is, if the variance in the gradual effects, σ^2_g, is equal to zero then $\rho = \beta\sigma^2_n / \beta\sigma^2_n = 1$.

We used a maximum likelihood framework to estimate ρ separately for each tree in the posterior distribution using the software BayesTraits [45]. Phylogenetic signal (λ) was estimated during the model-fitting process [52]. Default parameters were used throughout, and each tree was scaled to have a mean branch length equal to 0.1. The average correlation was then calculated across the posterior distribution.

(i) Estimating metastatic complexity

We characterized the level of 'metastatic complexity' within a given metastatic lineage by estimating the degree of phylogenetic dispersion, D [44], in the location of cells recovered from the primary tumour and distal site metastases. Metastatic lineages seeded by a single monophyletic branch with limited reseeding will have a lower degree of phylogenetic dispersion compared with lineages that have polyphyletic seeding and continued reseeding. We estimated D separately for each tree in the posterior distribution using the phylo.d function in the Caper R package [54]. The average D was then calculated across the posterior distribution.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. All data are available and referenced in the main text or the supplementary materials [55].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. G.B.: conceptualization, investigation, methodology, visualization, writing—original draft, writing—review and editing; S.R.A.: conceptualization, supervision, writing—review and editing; C.V.: conceptualization, methodology, supervision, writing—review and editing; K.J.P.: conceptualization, methodology, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. K.J.P. is a consultant for CUE Biopharma, Inc., and holds equity interest in CUE Biopharma, Inc., Keystone Biopharma, Inc., PEEL Therapeutics, Inc. and Kreftect Inc. S.R.A. holds equity interest in Keystone Biopharma, Inc. All other authors declare they have no competing interests.

Funding. US Department of Defense CDMRP/PCRP HT9425-23-1-0157 (GB) US Department of Defense CDMRP/PCRP W81XWH-20-10353, W81XWH-22–1-0680 (SRA) Prostate Cancer Foundation (GB, SRA, KJP) Patrick C. Walsh Prostate Cancer Research Fund (SRA) Leverhulme Trust Research Leadership Award RL-2019-012 (CV) The Commonwealth Foundation for Cancer Research (KJP)National Institute of Cancer PO1CA093900, U54CA210173, U01CA196390, and P50CA058236 (KJP).

Acknowledgements. We thank Robert Axelrod for helpful discussions about this work.

References

1. Greaves M, Maley CC. 2012 Clonal evolution in cancer. *Nature* **481**, 306–313. (doi:10.1038/nature10762)
2. Turajlic S, Sottoriva A, Graham T, Swanton C. 2019 Resolving genetic heterogeneity in *cancer*. *Nat. Rev. Genet.* **20**, 404–416. (doi:10.1038/s41576-019-0114-6)
3. Sottoriva A *et al*. 2015 A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **47**, 209–216. (doi:10.1038/ng.3214)
4. Jamal-Hanjani M *et al*. 2017 Tracking the Evolution of Non–Small-Cell Lung *Cancer*. *N. Engl. J. Med.* **376**, 2109–2121. (doi:10.1056/nejmoa1616288)
5. Spain L *et al*. 2023 Late-Stage Metastatic Melanoma Emerges through a Diversity of Evolutionary Pathways. *Cancer Discov.* 22–1427. (doi:10.1158/2159-8290.cd-22-1427)
6. Frankell AM *et al*. 2023 The evolution of lung cancer and impact of subclonal selection in TRACERx. *Nature* **616**, 525–533. (doi:10.1038/s41586-023-05783-5)
7. Househam J *et al*. 2022 Phenotypic plasticity and genetic control in colorectal cancer evolution. *Nature* **611**, 744–753. (doi:10.1038/s41586-022-05311-x)
8. Rogiers A, Lobon I, Spain L, Turajlic S. 2022 The Genetic Evolution of Metastasis. *Cancer Res.* **82**, 1849–1857. (doi:10.1158/0008-5472.can-21-3863)
9. Nguyen B *et al*. 2022 Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell* **185**, 563–575.(doi:10.1016/j.cell.2022.01.003)
10. Gui P, Bivona TG. 2022 Evolution of metastasis: new tools and insights. *Trends Cancer* **8**, 98–109. (doi:10.1016/j.trecan.2021.11.002)
11. Quinn JJ, Jones MG, Okimoto RA, Nanjo S, Chan MM, Yosef N, Bivona TG, Weissman JS. 2021 Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts. *Science* **371**. (doi:10.1126/science.abc1944)
12. Yang D *et al*. 2022 Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. *Cell* **185**, 1905–1923.(doi:10.1016/j.cell.2022.04.015)
13. Pagel M, Venditti C, Meade A. 2006 Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science* **314**, 119–121. (doi:10.1126/science.1129647)
14. Priestley P *et al*. 2019 Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **575**, 210–216. (doi:10.1038/s41586-019-1689-y)
15. Eldredge N, Gould SJ. 1972 Punctuated Equilibria: An alternative to phyletic gradualism. In *Models in paleobiology* (ed. TJM Schopf), pp. 82–115. San Francisco, CA: Cooper & Co.
16. Gould SJ, Eldredge N. 1977 Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* **3**, 115–151. (doi:10.1017/s0094837300005224)

17. Lenski RE, Travisano M. 1994 Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc. Natl Acad. Sci. USA* **91**, 6808–6814. (doi:10.1073/pnas.91.15.6808)
18. Elena SF, Cooper VS, Lenski RE. 1996 Punctuated evolution caused by selection of rare beneficial mutations. *Science* **272**, 1802–1804. (doi:10.1126/science.272.5269.1802)
19. Surya K, Gardner JD, Organ CL. 2023 Detecting punctuated evolution in SARS-CoV-2 over the first year of the pandemic. *Front. Virol* **3**, fviro.2023.1066147. (doi:10.3389/fviro.2023.1066147)
20. Baca SC *et al*. 2013 Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677. (doi:10.1016/j.cell.2013.03.021)
21. Gao R *et al*. 2016 Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat. Genet.* **48**, 1119–1130. (doi:10.1038/ng.3641)
22. Pagel M, Meade A. 2004 A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst. Biol.* **53**, 571–581. (doi:10.1080/10635150490468675)
23. Seidel S, Stadler T. 2022 TiDeTree: a Bayesian phylogenetic framework to estimate single-cell trees and population dynamic parameters from genetic lineage tracing data. *Proc. R. Soc. B* **289**, 20221844. (doi:10.1098/rspb.2022.1844)
24. Feng J, DeWitt III WS, McKenna A, Simon N, Willis AD, Matsen IV FA. 2021 Estimation of cell lineage trees by maximum-likelihood phylogenetics. *Ann. Appl. Stat.* **15**, 343–362. (doi:10.1214/20-aoas1400)
25. Kozlov A, Alves JM, Stamatakis A, Posada D. 2022 CellPhy: accurate and fast probabilistic inference of single-cell phylogenies from scDNA-seq data. *Genome Biol.* **23**, 37. (doi:10.1186/s13059-021-02583-w)
26. McKenna A, Gagnon JA. 2019 Recording development with single cell dynamic lineage tracing. *Development* **146**, dev169730. (doi:10.1242/dev.169730)
27. Pagel M, Lutzoni F. 2002 Accounting for phylogenetic uncertainty in comparative studies of evolution and adaptation. In *Biological evolution and statistical physics lecture notes in physics* (eds M Lässig, A Valleriani), pp. 148–161. Berlin, Heidelberg, Germany: Springer. (doi:10.1007/3-540-45692-9_8)
28. Webster AJ, Payne RJH, Pagel M. 2003 Molecular phylogenies link rates of evolution and speciation. *Science* **301**, 478–478. (doi:10.1126/science.1083202)
29. Venditti C, Meade A, Pagel M. 2006 Detecting the node-density artifact in phylogeny reconstruction. *Syst. Biol.* **55**, 637–643. (doi:10.1080/10635150600865567)
30. Field MG, Durante MA, Anbunathan H, Cai LZ, Decatur CL, Bowcock AM, Kurtenbach S, Harbour JW. 2018 Punctuated evolution of canonical genomic aberrations in uveal melanoma. *Nat. Commun.* **9**, 116. (doi:10.1038/s41467-017-02428-w)
31. Carson HL, Templeton AR. 1984 Genetic revolutions in relation to speciation phenomena: the founding of new populations. *Annu. Rev. Ecol. Syst.* **15**, 97–132. (doi:10.1146/annurev.es.15.110184.000525)
32. Rolland J *et al*. 2023 Conceptual and empirical bridges between micro- and macroevolution. *Nat. Ecol. Evol.* **7**, 1181–1193. (doi:10.1038/s41559-023-02116-7)
33. Venditti C, Pagel M. 2010 Speciation as an active force in promoting genetic evolution. *Trends Ecol. Evol.* **25**, 14–20. (doi:10.1016/j.tree.2009.06.010)
34. Reiter JG *et al*. 2020 Lymph node metastases develop through a wider evolutionary bottleneck than distant metastases. *Nat. Genet.* **52**, 692–700. (doi:10.1038/s41588-020-0633-2)
35. Wu Y *et al*. 2022 Spatiotemporal immune landscape of colorectal cancer liver metastasis at single-cell level. *Cancer Discov.* **12**, 134–153. (doi:10.1158/2159-8290.cd-21-0316)
36. Massagué J, Ganesh K. 2021 Metastasis-initiating cells and ecosystems. *Cancer Discov.* **11**, 971–994. (doi:10.1158/2159-8290.cd-21-0010)
37. Pagel M, Meade A, Barker D. 2004 Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* **53**, 673–684. (doi:10.1080/10635150490522232)
38. Lambert AW, Pattabiraman DR, Weinberg RA. 2017 Emerging biological principles of metastasis. *Cell* **168**, 670–691. (doi:10.1016/j.cell.2016.11.037)
39. de Groot AE, Roy S, Brown JS, Pienta KJ, Amend SR. 2017 Revisiting seed and soil: examining the primary tumor and cancer cell foraging in metastasis. *Mol. Cancer Res.* **15**, 361–370. (doi:10.1158/1541-7786.mcr-16-0436)
40. McPherson A *et al*. 2016 Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian *cancer*. *Nat. Genet.* **48**, 758–767. (doi:10.1038/ng.3573)
41. Brown D *et al*. 2017 Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations. *Nat. Commun.* **8**, 14944. (doi:10.1038/ncomms14944)
42. Al Bakir M *et al*. 2023 The evolution of non-small cell lung cancer metastases in TRACERx. *Nature* **616**, 534–542. (doi:10.1038/s41586-023-05729-x)
43. Gundem G *et al*. 2015 The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**, 353–357. (doi:10.1038/nature14347)
44. Fritz SA, Purvis A. 2010 Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conserv. Biol.* **24**, 1042–1051. (doi:10.1111/j.1523-1739.2010.01455.x)
45. Meade A, Pagel M. BayesTraits. See https://www.evolution.reading.ac.uk/BayesTraitsV4.1.1/BayesTraitsV4.1.1.html.
46. Jones MG, Khodaverdian A, Quinn JJ, Chan MM, Hussmann JA, Wang R, Xu C, Weissman JS, Yosef N. 2020 Inference of single-cell phylogenies from lineage tracing data using Cassiopeia. *Genome Biol.* **21**, 92. (doi:10.1186/s13059-020-02000-8)
47. McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J. 2016 Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907. (doi:10.1126/science.aaf7907)
48. Jones MG, Yang D, Weissman JS. 2023 New tools for lineage tracing in cancer in vivo. *Annu. Rev. Cancer Biol.* **7**, 111–129. (doi:10.1146/annurev-cancerbio-061421-123301)
49. Tavaré S. 1986 1984 Some probabilistic and statistical problems in the analysis of DNA sequences (ed. RM Miura). In *Eighteenth Annual Symposium on Some Mathematical Questions in Biology: DNA sequence analysis*, New York, NY, 28 May 1984, pp. 57–86. Providence, RI: American Mathematical Society.
50. Yang Z. 1994 Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Mol. Evol.* **39**, 306–314. (doi:10.1007/bf00160154)
51. Jombart T, Balloux F, Dray S. 2010 *adephylo* : new tools for investigating the phylogenetic signal in biological traits. *Bioinformatics* **26**, 1907–1909. (doi:10.1093/bioinformatics/btq292)
52. Pagel M. 1999 Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884. (doi:10.1038/44766)
53. Atkinson QD, Meade A, Venditti C, Greenhill SJ, Pagel M. 2008 Languages Evolve in Punctuational Bursts. *Science* **319**, 588. (doi:10.1126/science.1149683)
54. Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N, Pearse W. 2011 caper: Comparative Analyses of Phylogenetics and Evolution in R. (doi:10.32614/cran.package.caper)
55. Butler G, Amend SR, Venditti C, Pienta KJ. 2025 Supplementary material from: Punctuational evolution is pervasive in distal site metastatic colonization. Figshare (doi:10.6084/m9.figshare.c.7628654)