



Research

Cite this article: Mander L. 2016 A combinatorial approach to angiosperm pollen morphology. *Proc. R. Soc. B* **283**: 20162033. <http://dx.doi.org/10.1098/rspb.2016.2033>

Received: 15 September 2016

Accepted: 28 October 2016

Subject Areas:

palaeontology, plant science, taxonomy and systematics

Keywords:

plants, morphology, morphospace, combinatorics, pollen

Author for correspondence:

Luke Mander

e-mail: luke.mander@open.ac.uk

Electronic supplementary material is available online at <http://dx.doi.org/10.6084/m9.figshare.c.3575750>.

A combinatorial approach to angiosperm pollen morphology

Luke Mander

Department of Environment, Earth and Ecosystems, The Open University, Milton Keynes MK7 6AA, UK

LM, 0000-0003-4347-2705

Angiosperms (flowering plants) are strikingly diverse. This is clearly expressed in the morphology of their pollen grains, which are characterized by enormous variety in their shape and patterning. In this paper, I approach angiosperm pollen morphology from the perspective of enumerative combinatorics. This involves generating angiosperm pollen morphotypes by algorithmically combining character states and enumerating the results of these combinations. I use this approach to generate 3 643 200 pollen morphotypes, which I visualize using a parallel-coordinates plot. This represents a raw morphospace. To compare real-world and theoretical morphologies, I map the pollen of 1008 species of Neotropical angiosperms growing on Barro Colorado Island (BCI), Panama, onto this raw morphospace. This highlights that, in addition to their well-documented taxonomic diversity, Neotropical rainforests also represent an enormous reservoir of morphological diversity. Angiosperm pollen morphospace at BCI has been filled mostly by pollen morphotypes that are unique to single plant species. Repetition of pollen morphotypes among higher taxa at BCI reflects both constraint and convergence. This combinatorial approach to morphology addresses the complexity that results from large numbers of discrete character combinations and could be employed in any situation where organismal form can be captured by discrete morphological characters.

1. Introduction

Angiosperms (flowering plants) are an extremely diverse group of plants and are composed of an estimated 260 000 species [1]. Angiosperms are characterized by striking morphological diversity, and the group contains an array of life forms that includes herbs, epiphytes, bulbs, aquatic plants, shrubs and trees. Research into the morphological diversity of angiosperms has involved the development of morphospaces that allow the morphology of different taxa to be quantitatively compared [2]. For example, morphospaces for angiosperm flowers have been used to investigate the frequency with which certain flower morphologies occur in nature [3] and also to examine the disparity of floral shapes [4]. Studies of fossil angiosperm pollen grains have shown that the taxonomic diversity and morphological disparity of angiosperms were decoupled during the evolutionary radiation of this clade during the Cretaceous period (approx. 150–65 Ma) [5], and recent work has highlighted that plant clades, including the angiosperms, typically attain a high level of morphological disparity early in their evolutionary history [2].

In general, researchers recognize two approaches to the construction of a morphospace. The first approach is theoretical, which involves using a mathematical model to generate morphologies and may also include comparison with real-world forms [6,7]. In this context, simulations of plant phenotypes [8] are a botanical counterpoint to simulations of animal shell shape [9] and are a classic example of theoretical morphospace construction. The second approach is empirical, which involves constructing a morphospace from a set of empirical morphological data [2]. In an empirical morphospace, the form of an organism is typically represented by a number of discrete characters, and there is often overlap between these characters and the set of characters

that could be used for the purpose of classification or phylogenetic analysis [10]. The dimensionality of an empirical morphospace is usually reduced using multivariate ordination [6,7], and the morphological disparity of a clade can be tracked through time using a distance metric such as mean pairwise dissimilarity (e.g. [5,10]). The morphospaces for plants constructed by Lupia [5] and Oyston *et al.* [2] are empirical, and in this respect they are similar to foundational studies on the morphological diversity of marine animals that are also based on discrete characters (e.g. [10,11]).

In addition to these theoretical and empirical approaches, however, there are also raw morphospaces, which are formulated in terms of observed morphological variation but prior to any multivariate ordination [7]. These morphospaces are empirical in the sense that aspects of real form are captured by the enumeration of discrete variables, rather than using a generative model, but they are also theoretical in the sense that they can represent forms that do not exist in nature [7,12]. Raw morphospaces are useful tools in the investigation of organic form because they facilitate ‘the study of the evolution of the actual in the realm of the possible’ [7, p. 10]. Stebbins produced a raw morphospace for angiosperm flowers by constructing a grid measuring 16 cells by 16 cells [3]. Each of the 256 cells in this grid represented one of 256 possible combinations of primitive versus advanced conditions for the eight binary floral characters that Stebbins analysed [3]. This grid allowed Stebbins to graphically represent morphological types produced by different discrete character combinations, and to demonstrate that the proportion of floral shapes that exist in nature is a limited subset of the floral shapes that might be structurally possible [3]. For Stebbins, one of the primary problems with his analysis was ‘the distortion produced by the recognition of only two conditions in respect to each character. This [he noted] was necessary because of the great complexity which would have resulted if more conditions had been recognized’ [3, p. 302].

In this paper, I address the complexity that results from large numbers of discrete character combinations by approaching angiosperm pollen morphology from the perspective of enumerative combinatorics. I begin with a schematic example that demonstrates the construction of a raw morphospace for angiosperm pollen. I use this approach to study angiosperm pollen morphology in the hyperdiverse Neotropics, and this involves comparing the morphology of 1008 real-world pollen grains to 3 643 200 algorithmically generated pollen morphotypes. The approach I outline is not restricted to angiosperm pollen and could be used in any situation where organismal form can be captured by discrete morphological characters.

2. A combinatorial approach to angiosperm pollen morphology

The overall morphology of a single angiosperm pollen grain is determined by the nature and arrangement of several different individual characters. The states of these characters are typically described using specialist terminology, which allows researchers to communicate information about pollen (e.g. [13]). This specialist terminology treats each of these character states as discrete elements that contribute to the overall morphology of a pollen grain. Different pollen

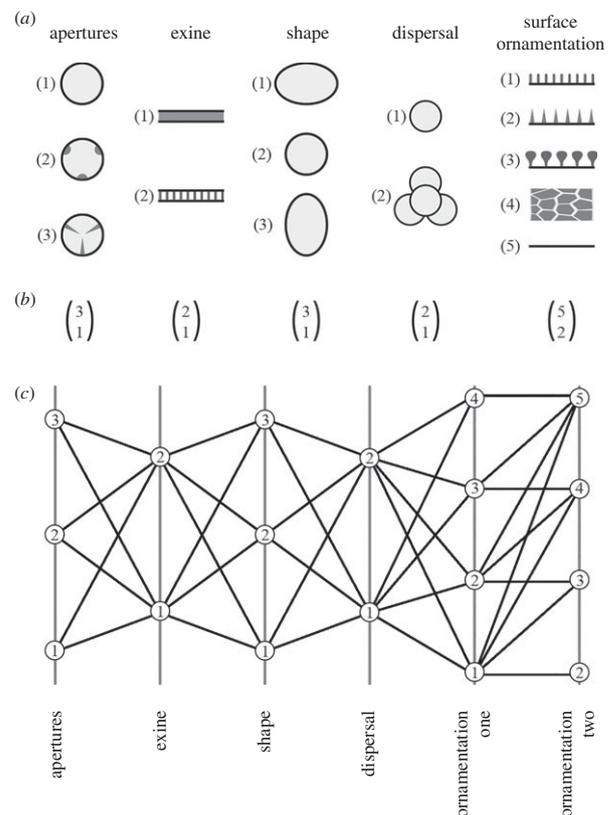


Figure 1. (a) Schematic illustration of characters that make up an angiosperm pollen grain. Aperture character states: inaperturate (1), triporate (2), tricolpate (3). Exine character states: atectate (1), tectate (2). Shape character states: oblate (1), spheroidal (2), prolate (3). Dispersal character states: monad (1), tetrad (2). Surface ornamentation character states: baculate (1), echinate (2), clavate (3), reticulate (4), no ornamentation (5). Different pollen morphotypes reflect different combinations of these character states. (b) Binomial coefficients for each character. (c) Diagram showing how these different combinations can be represented using a parallel-coordinates plot.

morphotypes reflect different combinations of these character states. This is shown in figure 1a, which contains five morphological characters, each with a number of different states, which can be used to construct a schematic pollen grain (descriptive terminology in this paper follows [13]).

(a) Counting the number of possible angiosperm pollen morphotypes

One possible combination of the character states shown in figure 1a is: 3, 2, 2, 1, 2, 4. In the language of formal descriptive palynology [13], this combination would produce a tricolpate pollen grain with a tectate exine, spheroidal shape, monad dispersal and an echinate–reticulate surface ornamentation. In more general terminology, this pollen grain has three elongated apertures in its exine, which is the outer coating of a pollen grain that is composed of sporopollenin [13] and in this case possesses a distinctive layer known as the tectum, which forms a roof over its internal structures [13]. This pollen grain has a spherical three-dimensional shape, is dispersed at maturity by its parent plant as a single grain (a monad) and has spines longer than 1 μm together with a network-like pattern on the surface of its exine. Generating this combination involves selecting one state from each list of aperture, exine, shape and dispersal character states. Some pollen grains possess just a single

type of surface ornamentation, but others have two distinct types that form primary and secondary surface ornamentation (although the two are not hierarchically related). *Razisea spicata* (Acanthaceae), for example, has sparse echini and dense baculae (see [13] for terminology). Accordingly, this combination involves selecting two states from the list of surface ornamentation character states.

I have thought of this process of character state selection as a combinatorial counting problem, in which the number of ways of selecting k objects from a list of n objects, if the order of selection is irrelevant, is given by

$$\binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!} & \text{if } 0 \leq k \leq n, \\ 0 & \text{otherwise.} \end{cases}$$

The numbers $\binom{n}{k}$, which can be read ‘ n choose k ’, are the binomial coefficients, and the value of $\binom{n}{k}$ is 0 when $k < 0$ or $k > n$ because there are no ways to select fewer than zero or more than n objects from a list of n objects [14]. In the schematic example shown in figure 1, the values of $\binom{n}{k}$ for each of the five characters are shown in figure 1*b*.

For the aperture, exine, shape and dispersal characters, only one state is selected from each character state list, and the binomial coefficients for these characters are written as $\binom{n}{1}$ (figure 1*b*). For the aperture character, for example, there are $\binom{3}{1} = 3$ ways to select a single character state from this list. For surface ornamentation, two states are chosen from the list of five character states, and the binomial coefficients for this character are written as $\binom{5}{2}$ (figure 1*b*).

In this case, there are $\binom{5}{2} = 10$ ways to select two states from this list. By the product rule, the number of all possible combinations of these character states is

$$\binom{3}{1} \binom{2}{1} \binom{3}{1} \binom{2}{1} \binom{5}{2} = 360.$$

In other words, there are 360 unique pollen morphotypes that can be generated from the five characters shown in figure 1*a* using n choose k combination.

(b) Enumerating and displaying angiosperm pollen morphotypes

For small values of $\binom{n}{k}$, it is possible to count and enumerate combinations manually. For example, there are six combinations of two letters that can be produced from the four letters ABCD, and this is given by $\binom{4}{2}$ or 4 choose 2.

This can be calculated using factorials or by using Pascal’s triangle for binomial coefficients, and these six combinations can be enumerated manually by hand: AB AC AD BC BD CD. However, manual counting and enumerating becomes rapidly intractable as the values of $\binom{n}{k}$ increase. To overcome this, I have written an algorithm using the Python programming language that generates angiosperm pollen morphotypes by n choose k combination of morphological

characters. An example of this algorithm, which counts and enumerates all possible combinations of the character states shown in figure 1*a*, is provided in the electronic supplementary material. This algorithm takes lists of morphological characters as its input, enumerates the combinations of these morphological characters using n choose k combination and then writes these combinations to a .csv file with headers for each character. An example of a .csv file containing all possible combinations of the character states shown in figure 1*a* is provided in the electronic supplementary material.

I have visualized these character combinations using a parallel-coordinates plot (figure 1*c*). In this plot, each vertical axis represents a different morphological character and lines visualize associations between character states. Parallel-coordinates plots can be thought of as multipartite graphs, with each vertical axis corresponding to a graph partition [15], and they have been used to visualize large volumes of network traffic data [16]. This plot represents a raw morphospace that is bounded and contains a range of possible morphologies (figure 1*c*). Provided that the same characters and states are used, real-world pollen grains can be mapped onto this morphospace.

3. Hyperdiversity in the Neotropics

I have applied this combinatorial approach to the hyperdiverse angiosperm-dominated tropical rainforests of South America. This region of the world contains the highest levels of plant diversity on the Earth and supports an estimated approximately 16 000 tree species alone [17]. When both extant and extinct taxa are considered, the total diversity of all Neotropical plant species that have ever existed is potentially enormous. I have focused on the morphology of angiosperm pollen grains produced by plants growing on Barro Colorado Island (BCI). Barro Colorado Island is a 1560 ha island situated in the Panama Canal that supports hyperdiverse lowland moist tropical forest, and a 50 ha plot on BCI that has been monitored since 1980 contains just over 300 tree and shrub species [18]. This region is of particular interest in the context of this paper because such high species diversity offers an opportunity to confront the challenge of analysing large numbers of discrete character combinations [3]. The vegetation of BCI also represents a single well-studied flora with which to examine the patterns of morphospace occupation by different plant groups, and to explore the properties of occupied and unoccupied morphospace (e.g. [12]).

(a) Morphospace construction

I used the methods outlined in figure 1 to construct a raw angiosperm pollen morphospace based on 60 aperture character states, 4 exine character states, 10 shape character states, 6 dispersal character states and 23 surface ornamentation character states (a full list of character states is provided in the electronic supplementary material). For surface ornamentation, two states are chosen from the list of 23 character states, and there are $\binom{23}{2} = 253$ ways to select two states from this list. By the product rule, the number of all possible combinations of these character states is

$$\binom{60}{1} \binom{4}{1} \binom{10}{1} \binom{6}{1} \binom{23}{2} = 3\,643\,200.$$

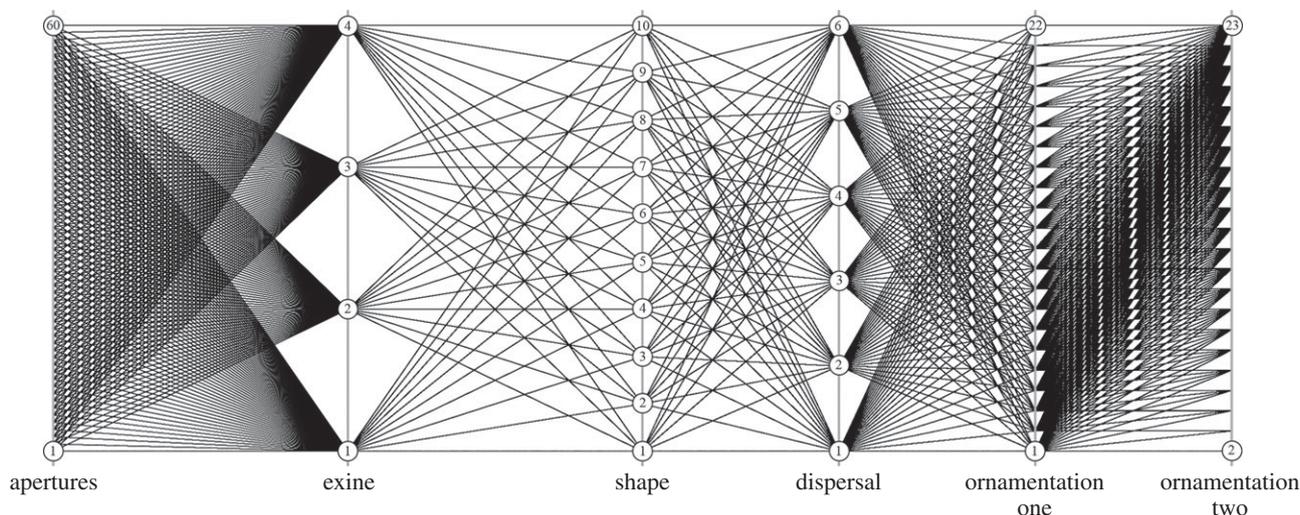


Figure 2. Parallel-coordinates plot of 3 643 200 angiosperm pollen morphotypes generated algorithmically by n choose k combination. Numbers enclosed within white circles represent each of the character states used to produce this morphospace. Only the first and last aperture and surface ornamentation character states are labelled to prevent overcrowding of these axes. A full list of character states corresponding to these numbers is provided in the electronic supplementary material.

The raw morphospace that represents these 3 643 200 algorithmically generated pollen morphotypes is shown in figure 2, and this allows associations between character states to be visualized. These character states were selected from the glossary of Punt *et al.* [13], and each is known to exist in nature. Some combinations of these character states are also known to exist in nature. For example, the combination aperture 3, exine 4, shape 9, dispersal 1, ornamentation one 2, ornamentation two 23 produces a pollen grain with a single annulate pore, a tectate exine, spheroidal shape, monad dispersal and scabrate surface ornamentation. This combination represents an archetypal grass pollen grain [19]. However, such real-world morphologies are a limited subset of the total 3 643 200 possible combinations in this morphospace, and consequently some of the morphologies shown in figure 2 are theoretical and not known to exist in nature.

(b) Patterns of morphospace occupation by real-world Neotropical angiosperm pollen grains

To map real-world Neotropical angiosperm pollen grains onto this morphospace, I examined a monograph containing morphological descriptions and images of pollen grains produced by plants growing on BCI [20]. I scored the angiosperm pollen grains in this monograph for each of the characters used to construct figure 2. I excluded the following taxa from this analysis: (i) those lacking a plate, (ii) those with missing characters and (iii) polyads. I followed the higher taxonomic classifications in [20], and this resulted in a dataset consisting of 115 angiosperm plant families, which contain a total of 1008 species and produce a total of 468 pollen morphotypes (table 1; a full list of taxa and character scorings is provided in the electronic supplementary material). This is just 0.01% of the 3 643 200 pollen morphotypes that were generated algorithmically by n choose k combination and used to construct the raw morphospace shown in figure 2.

I split these 115 families into monocots (monocotyledons: angiosperms whose seeds contain a single cotyledon) and dicots (dicotyledons: angiosperms whose seeds contain two

cotyledons). This partitioned dataset consists of 19 monocot families, which contain a total of 184 species and produce 68 pollen morphotypes, and 96 dicot families, which contain a total of 824 species and produce 407 pollen morphotypes (table 1). I then mapped each BCI monocot and dicot pollen morphotype onto the raw morphospace shown in figure 2. This shows how the morphospace occupied by the pollen of these two plant groups compares with the morphospace created by 3 643 200 algorithmically generated pollen morphotypes (figure 3).

The number of real-world BCI morphotypes in this analysis is clearly too few to occupy the entire raw morphospace. However, visual inspection of figure 3 highlights that, despite this, a considerable amount of the available morphospace is occupied by the BCI pollen morphotypes (figure 3). This is particularly notable in the morphospace occupied by dicot pollen, and for this plant group there are some associations between character states that are completely saturated. In this context, saturation may be defined as the proportion of possible associations between different character states that are realized in nature. For example, there is an association between the tectate exine character state and each of the shape character states, and there is also an association between the monad dispersal state and each of the shape character states (figure 3*b*). Both of these examples therefore have a saturation score of 1. There is also a strong, but not saturated, association between these character states for monocot pollen (figure 3*a*). For example, there are 10 possible associations between the tectate exine character state and the shape character states, and six of these are realized in the BCI dataset of monocot pollen (figure 3*a*). This gives a saturation score of 0.6.

Both monocots and dicots produce pollen with a variety of surface ornamentation patterns, and for both plant groups this area of the morphospace is well explored (see the ornamentation one axis of figure 3). For both of these plant groups, there is a strong association between the ornamentation one character states and the uppermost extremity of the ornamentation two axis (figure 3). This association represents pollen grains that have just a single type of surface ornamentation. Associations between character states on the

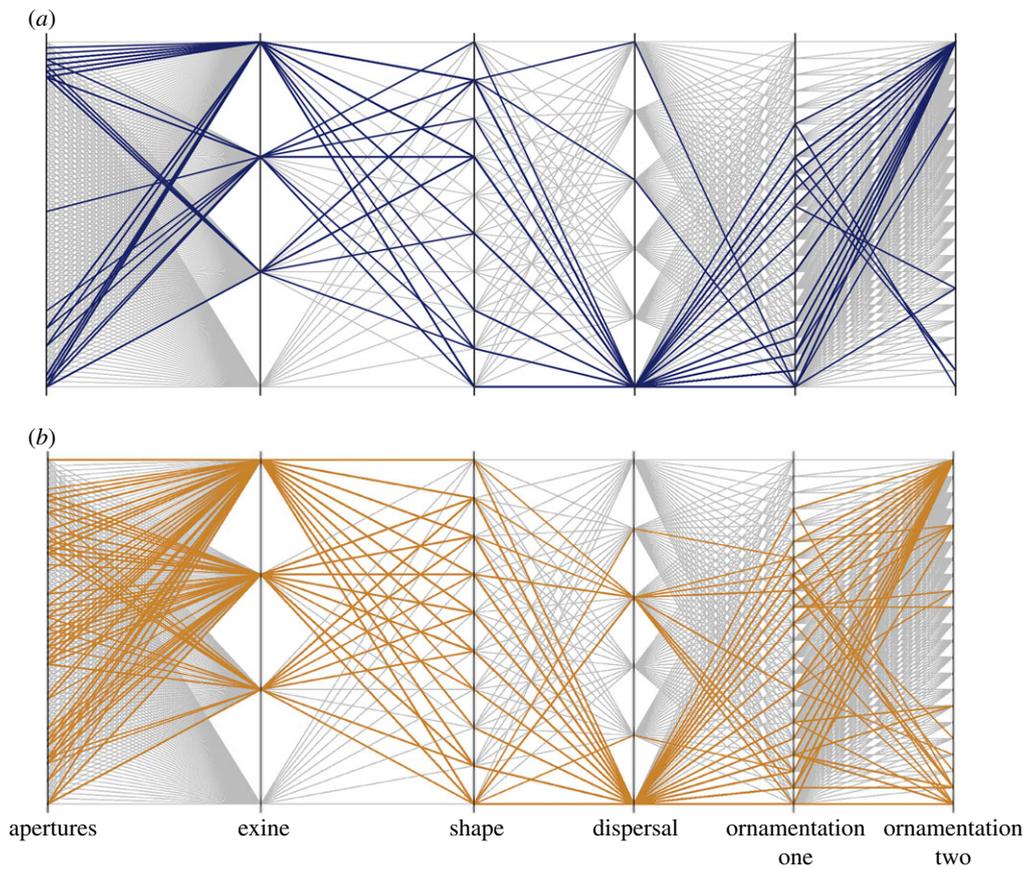


Figure 3. Parallel-coordinates plots showing how pollen grains produced by angiosperms growing on Barro Colorado Island map onto the morphospace shown in figure 2. In both plots, this morphospace is shown in light grey and is overlain by (a) monocot pollen morphotypes in dark blue and (b) dicot pollen morphotypes in orange. A full list of taxa and character scorings is provided in the electronic supplementary material.

Table 1. Summary of the data and results discussed in this paper.

	families (<i>n</i>)	species (<i>n</i>)	morphotypes (<i>n</i>)	morphotypes as a proportion of species
<i>n</i> choose <i>k</i> morphospace	—	—	3 643 200	—
BCI angiosperm pollen	115	1008	468	0.46
BCI monocots	19	184	68	0.37
BCI dicots	96	824	407	0.49
BCI Araceae (monocot)	—	39	22	0.56
BCI Poaceae (monocot)	—	53	6	0.11
BCI Papilionoideae (dicot)	—	60	47	0.78
BCI Rubiaceae (dicot)	—	67	45	0.67

ornamentation one axis and other character states on the ornamentation two axis represent pollen grains that have both primary and secondary surface ornamentation. Both monocots and dicots have evolved the means to produce pollen with both primary and secondary ornamentation, and this morphospace shows how this capability has expanded their morphological variety (figure 3).

This mapping also shows that the pollen grains produced by BCI monocots and dicots occupy strikingly different regions of morphospace (figure 3). This is highlighted by the differences in monocot and dicot aperture character states. Apertures are a key character in the study of angiosperm pollen, and are thought to be a key difference between the pollen of the monocot and dicot clades [21].

The aperture character states of BCI monocots are situated in two groups that are positioned at opposite poles of the aperture character axis (figure 3a). The group at the lower extremity of the aperture character axis contains inaperturate, monoporate, triporate, pantoporate and stephanoporate morphotypes, while the group at the upper extremity contains zonorate, trichotomosulcate, disulcate and monosulcate morphotypes (see [13] for terminology). These two groups are separated by a region of morphospace that is largely unoccupied by monocot morphotypes (figure 3a). This region contains tricolpate, stephanocolpate and tricolporate forms (see [13] for terminology) and is occupied by a variety of dicot pollen morphotypes (figure 3b). These aperture types are highly characteristic of the eudicot

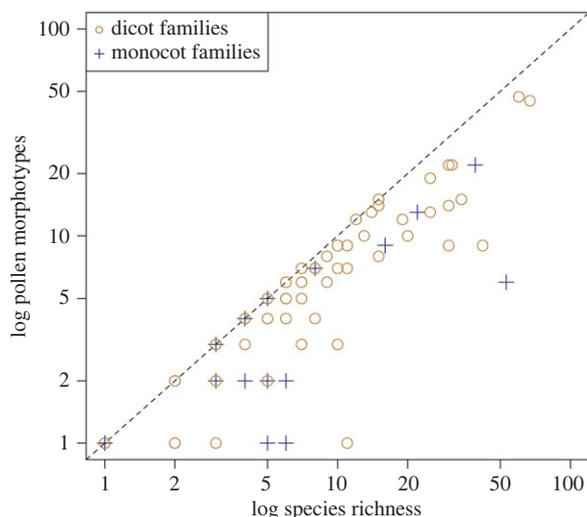


Figure 4. Scatter plot comparing the species richness of angiosperm families from Barro Colorado Island with the number of pollen morphotypes produced by each of these families. Data points have been plotted for each of the 19 monocot and 96 dicot families analysed in this paper, but data points for certain families overlap. Both axes are plotted on a logarithmic scale. Diagonal dashed line represents a line of equality. (Online version in colour.)

plant clade, but the selective pressure (if any) driving their adoption by this plant group is largely unknown [22].

4. Discussion

(a) Pollen grains and Neotropical plant diversity

There are fewer pollen morphotypes than plant species in the dataset of BCI angiosperms that I have used in this paper. This is reflected in the number of pollen morphotypes expressed as a proportion of species for all angiosperms in the dataset as well as monocots and dicots (table 1). Some of the plant families that constitute the BCI flora produce very few pollen morphotypes. Such families include the Poaceae (grasses), which are taxonomically very rich (53 species in this BCI dataset) but only produce six pollen morphotypes (table 1). Other families are also very speciose but instead produce a large number of pollen morphotypes; an example is the Papilionoideae (a subfamily of legumes), which contains 60 species in this BCI dataset and produces 47 pollen morphotypes (table 1).

This pattern, in which a plant group is characterized by the production of fewer pollen morphotypes than species, has been frequently observed in many geographical regions and time periods, and pollen grains are said to suffer from low taxonomic resolution [23]. This pattern holds for the BCI flora as a whole, and for certain families (table 1), but when plant species richness is plotted against the number of pollen morphotypes for all families at BCI, there is considerable scatter in the distribution of data points (figure 4). This shows that the relationship between the taxonomic diversity of the BCI flora and the number of pollen morphotypes produced by its constituent plants is too variable to be encompassed by a single notion of low taxonomic resolution.

Some data points are indeed situated well below the line of equality, and pollen morphotypes clearly underestimate the species diversity of these families (figure 4). However, there are 32 families that plot directly on the line of equality and 23 families that miss the line by just 1 (both excluding

families represented by a single species on BCI), and there are several other families that plot close to this line of equality (figure 4). For these families, assemblages of dispersed pollen grains will provide an accurate or at least reasonable reflection of their species diversity (figure 4). Additionally, there are some plant species on BCI that produce dimorphic pollen; examples include *Cocos nucifera* (the coconut tree), which produces pollen with either a monosulcate or a trichotomosulcate aperture. Indeed, among monocots there are several families in which pollen aperture type is not consistent within anthers, and these include the Agavaceae and Iridaceae (both Asparagales), as well as the Arecaceae, which contains *C. nucifera* [21]. It is possible, therefore, that in certain circumstances data points could lie above the line of equality in figure 4, and that assemblages of dispersed pollen grains could overestimate the diversity of the source vegetation. The patterns of morphospace occupation by BCI angiosperm pollen grains (figure 3) therefore reflect a situation in which some plant families, such as the Papilionoideae, contribute a large quantity of morphological variety, whereas other taxa, such as the Poaceae, contribute much less morphological variety.

(b) The structure and utility of a simplified palynological taxonomy

In this paper, the morphology of angiosperm pollen grains has been encoded in a system of discrete characters (see figure 2 and the electronic supplementary material). This system contains fewer characters than would be employed by a palynologist seeking to classify pollen grains for the purpose of reconstructing vegetation history, and omits certain characters that may be phylogenetically informative such as the composition of the endexine and the nature of the intine [21]. Additionally, when observed using a microscope, the surface ornamentation of pollen grains can appear to vary continuously between individuals and species in a manner that is not well captured by the discrete approach that is taken here. As such, both the total number of character combinations that are represented by the raw morphospace shown in figure 2, and the number of BCI pollen morphotypes mapped onto this morphospace in figure 3, represent minimum diversity. The inclusion of just two size classes $\binom{2}{1}$, for example, would double the total number of possible character state combinations used to construct the raw morphospace shown in figure 2:

$$\binom{60}{1} \binom{4}{1} \binom{10}{1} \binom{6}{1} \binom{23}{2} \binom{2}{1} = 7\,286\,400.$$

By contrast, the removal of secondary ornamentation would substantially reduce the total number of possible character state combinations:

$$\binom{60}{1} \binom{4}{1} \binom{10}{1} \binom{6}{1} \binom{22}{1} = 316\,800.$$

Secondary ornamentation is a rare character in the BCI flora (there are just 65 BCI pollen morphotypes with secondary ornamentation), and in this respect BCI is typical of angiosperm pollen morphology in other floras.

This discrete character-based system of encoding pollen morphology is comparable with formal taxonomic

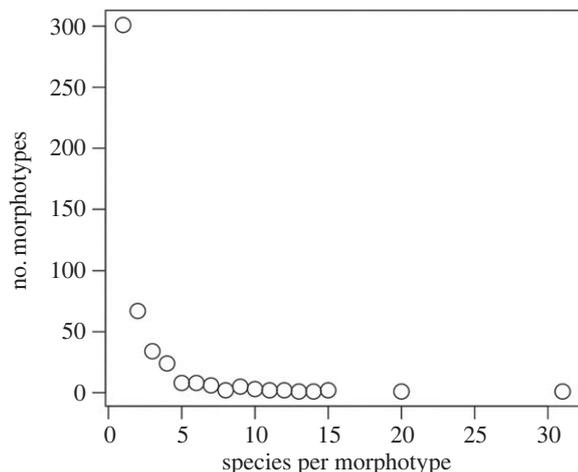


Figure 5. Scatter plot showing the hollow-curve relationship between the number of species per angiosperm pollen morphotype and the number of angiosperm pollen morphotypes on Barro Colorado Island.

descriptions of pollen grains in the sense that both involve breaking pollen morphology down into individual characters, which are then described using specialist terminology (figure 1*a*) [13]. Furthermore, in spite of its simplicity, this system contains a relationship between the number of species per pollen morphotype and the number of pollen morphotypes that is essentially of the same mathematical form as the relationship between the number of species per genus and the number of genera in other taxonomic systems (figure 5). These include classifications of plants, sponges, molluscs, fishes, beetles and birds [24,25]. The correspondence between figure 5 and similar plots from other taxonomic systems [24,25], which are often presented on log-log plots, suggests that studying the structure of palynological taxonomy through evolutionary time could provide data on the diversification history of plant clades [26].

Together with data comparing plant species richness and the number of pollen morphotypes for all families at BCI (figure 4), figure 5 indicates that angiosperm pollen morphospace at BCI (figure 3) has been filled mostly by pollen morphotypes that are unique to single species (301 instances; top-left data point in figure 5). There are 167 pollen morphotypes that are produced either by different species within a single family, such as the archetypal grass pollen grain, or by taxa within two or more families, and these repetitions represent 36% of the pollen morphotypes at BCI (electronic supplementary material, figure S1). This emphasizes that when studying assemblages of Neotropical angiosperm pollen grains, even a highly simplified system consisting of the five discrete characters used in this paper can capture reasonable data on the diversity of the source vegetation. The addition of further characters would be likely to increase the number of pollen morphotypes and move the data points in figure 5 upwards and to the left. If these results are applicable to other regions of the world and through geological time, then they provide a degree of support for studies that use pollen grains to reconstruct the diversity and composition of tropical vegetation through time.

(c) The production and distribution of biological form

In this paper, I have not generated morphologies using a growth model that incorporates morphogenetic processes

(cf. [6,27,28]). Instead, I have represented morphogenesis in angiosperm pollen by the algorithmic combination and enumeration of discrete morphological characters (e.g. figure 1), and this combinatorial approach allows for the explicit comparison of occupied and unoccupied morphospace. Each of the 3 643 200 morphotypes that can be produced from the characters I have used in this study (figure 2 and table 1; electronic supplementary material) is biologically plausible and structurally possible. There are no morphotypes with both tricolpate and monocolpate apertures, for example, and there are no morphotypes that contain a character state that is not known to exist in nature. However, the number of these plausible and possible angiosperm pollen morphotypes is far greater than the number of angiosperm species on Earth as a whole (approx. 260 000 [1]) and in the BCI dataset I have used here (1008; table 1). This means that when real-world forms are mapped onto the raw morphospace that contains these 3 643 200 morphotypes, there are large areas of unoccupied morphospace (figure 3). Although the BCI dataset used in this study represents just a small fraction of angiosperm life, I suspect that a similar analysis incorporating all known angiosperm species would produce a similar pattern.

One way of interpreting patterns of morphospace occupation is to think of the entire morphospace as a fitness landscape, with occupied areas representing local optima and unoccupied areas representing less fit alternatives [12]. This view of morphospace occupation is informed by Wright's studies of allele combinations and fitness landscapes, in which he viewed a possible 10^{1000} allele combinations as a rugged field with 'an enormous number of widely separated harmonious combinations' [29, p. 358]. Wright viewed these harmonious combinations as adaptive peaks separated by maladaptive valleys, and Stebbins, without directly referencing Wright's work, seems to interpret the patterns in his floral morphospace in a similar manner: 'there are three types of combinations which, although structurally possible, are found so rarely that they are apparently poor adaptations, usually of low survival value' [3, p. 314].

Many attempts have been made to establish the function of the morphological characters that together make up a pollen grain (e.g. [30]), and discussion has ranged from the role of surface ornamentation in pollination [22] to the role of apertures in the desiccation, hydration and germination of pollen grains [31]. Consequently, it is currently unclear which morphological features of pollen grains are adaptive and which are not, and this means that the relative roles of chance versus necessity in generating the morphological diversity of pollen grains are largely unknown. Indeed, the wide range of plausible functional roles for the morphological features of pollen grains suggests that they are 'not "optimally" designed for a specific function, but merely structures that work with varying efficiency in a specific ecological and evolutionary context' [30, p. 182]. With this in mind, I do not interpret the areas of unoccupied morphospace in figure 3 as regions containing pollen morphotypes that are maladaptive in the context of the BCI flora. It is possible, however, that further experimental work on the function of specific morphological features may reveal regions of morphospace that are characterized by forms with specific functional properties. Such work could also include analyses of the tendency of certain morphological features to arise by self-assembly processes, which may produce forms that

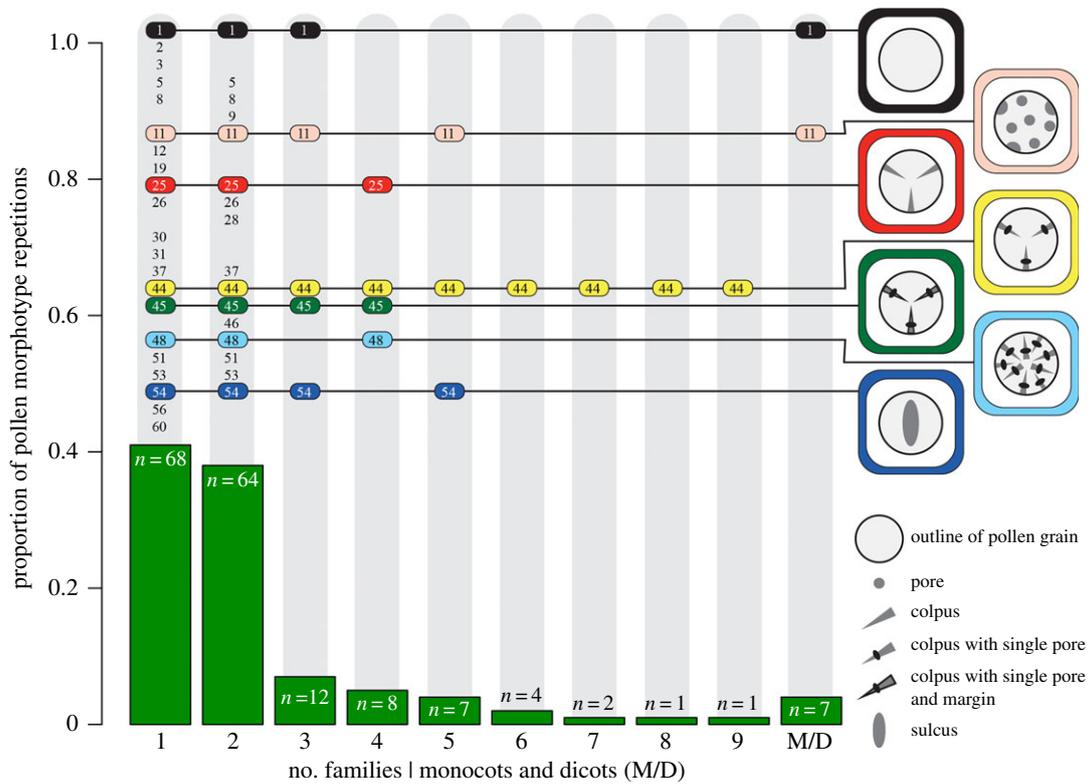


Figure 6. Proportions of angiosperm pollen morphotypes that are repeated among higher taxa at BCI (green vertical bars). Numbers enclosed within grey vertical boxes show the distribution of aperture character states that are repeated among higher taxa at BCI (see the electronic supplementary material, dataset S1 for a full list of aperture character states). Key aperture character states repeated among three to nine families as follows: inaperturate (1), pantoporate (11), tricolpate (25), tricolporate (44), tricolporate marginate (45), pantocolporate (48), monosulcate (54). Cartoons to the right are schematic pollen grains that graphically display each of these aperture character states.

reflect convergence upon a minimum free energy state rather than the effects of natural selection [32].

Alternative interpretations of morphospace occupation include the suggestion that the present-day distribution of morphology (e.g. figure 3) reflects an earlier more continuous distribution that has been subsequently fragmented by extinction [12], as well as Gould's classic idea that chance played a major role in distributing form, with the contingent elimination of forms early in their evolutionary history producing a clumpy distribution of morphology among living organisms [12,33]. The forms of such living organisms may represent relatively small historically defined groups that are drawn from more or less unlimited possibilities (e.g. [3]), or could instead be fundamentally limited by structural and/or phylogenetic constraints (see [5] for angiosperm pollen and [34] for animal skeletons). Results of this analysis suggest that areas in morphospace that are unoccupied by modern organisms (e.g. figure 3) could reflect the fact that many potentially viable morphologies have not yet been produced by the evolutionary process [35].

However, if a morphospace is defined using character states that are sufficiently broad in scope (e.g. [34]), then the possibility of unrealized morphologies can be effectively eliminated, whereas if large numbers of character states are used to define a morphospace (as in figure 2), then the number of unrealized morphologies can be increased at will. Despite this scaling issue, each of the character states used to construct the raw morphospace shown in figure 2 is known to exist in nature and is very general compared with the complexity of pollen morphology as viewed through a microscope [13]. Additionally, none of the 3 643 200 pollen morphotypes that

these general and simplified character states produce seems biologically implausible. In this light, the possibility of large numbers of unrealized angiosperm pollen morphologies seems real, and the number of such potential forms that have not yet evolved appears to be vast. This is in contrast to analyses of animal skeletons, for which 'the number of potential skeletal designs, defined in very general terms, is not inordinately large' [34, p. 350]. However, the difference in the outcomes of these two analyses (angiosperm pollen versus animal skeletons) is primarily a consequence of the different scope of the characters that are defined in each case.

In the dataset of BCI angiosperm pollen I have analysed in this paper, there are 167 morphotypes that are repeated among higher taxa (electronic supplementary material, figure S1; figure 6), and these repetitions highlight the numbers and kinds of morphotypes that reflect phylogenetic constraint as well as evolutionary convergence. The greatest number of these repetitions occurs among different species or genera within single families ($n = 68$; figure 6). Examples of this include the pollen of grasses (single annulate pore, tectate exine, spheroidal shape, monad dispersal, scabrate surface ornamentation; 31 repetitions in one family), laurels (inaperturate, intectate exine, spheroidal shape, monad dispersal, echinate surface ornamentation; 11 repetitions in one family) and figs (diporate, tectate exine, oblate shape, monad dispersal, psilate surface ornamentation; 7 repetitions in one family). These morphologies exhibit strong phylogenetic constraint because they are common within these families but are restricted to them.

By contrast, there are relatively few morphotypes that are repeated in three to nine families (figure 6). These 35

morphotypes are notably similar in terms of their aperture character state: 23 of these morphotypes share the same aperture character state (tricolporate; 65.7%), and all the morphotypes that are repeated among six to nine families share this character state (figure 6). With respect to aperture character states, this highlights a degree of nestedness in the distribution of angiosperm pollen morphology at BCI. Seven different morphotypes are repeated in both the monocot and dicot clades (figure 6). These morphotypes also have a very uniform set of aperture characters: two are inaperturate and five have multiple pores (pantoporate; figure 6). As quoted in [22], ‘The systematic distribution of periporate [pantoporate of many authors] pollen in extant angiosperms indicates that this aperture configuration evolved independently in many different groups’ [36, p. 406]. The repetition of these aperture configurations across a fundamental divide in plant phylogeny is an example of evolutionary convergence, and in this respect, the BCI flora reflects a broader pattern in angiosperm evolution.

5. Concluding remarks

The combinatorial approach to angiosperm pollen morphology that I have outlined in this paper provides an example of how the general challenge of analysing large numbers of discrete character combinations [3] can be overcome. This approach generates biological forms by algorithmically combining character states and enumerating the results of these combinations (figure 1). I have used this approach to generate 3 643 200 angiosperm pollen morphotypes and produce a raw morphospace (figure 2). Some of these morphotypes correspond to real-world forms, whereas others do not.

I have mapped the pollen of 1008 species of Neotropical angiosperms growing on BCI, Panama, onto this raw

morphospace (figure 3). This allows realized and unrealized forms to be compared and shows that despite the relatively small size of the BCI dataset, these real-world pollen grains occupy a considerable amount of the raw morphospace (figure 3). This highlights that, in addition to their well-documented taxonomic diversity, Neotropical rainforests also represent an enormous reservoir of morphological diversity. Some plant families, such as the Papilionoideae, contribute a large amount of morphological diversity, whereas other taxa, such as the Poaceae, contribute much less morphological diversity (figure 4).

In this paper, the morphology of angiosperm pollen grains has been encoded in a highly simplified system of five discrete characters. Despite this simplicity, this system contains a relationship between the number of species per pollen morphotype and the number of pollen morphotypes that has essentially the same mathematical form as the relationship between the number of species per genus and the number of genera in other taxonomic systems (figure 5). This highlights that the angiosperm pollen morphospace at BCI (figure 3) has been filled mostly by pollen morphotypes that are unique to single species (figure 5). Angiosperm pollen morphotypes that are repeated among higher taxa at BCI reflect both phylogenetic constraint and evolutionary convergence (figure 6).

Data accessibility. The datasets and code supporting this article have been uploaded as part of the electronic supplementary material.

Competing interests. I declare I have no competing interests.

Funding. I received no funding for this work.

Acknowledgements. I am grateful to Claire Belcher and Carlos Jaramillo, who provided useful comments on an earlier version of this work. I am grateful to the reviewers of this work, particularly Roger Thomas, whose thoughtful suggestions clarified and improved the ideas presented here.

References

- Soltis PS, Soltis DE. 2004 The origin and diversification of angiosperms. *Am. J. Bot.* **91**, 1614–1626. (doi:10.3732/ajb.91.10.1614)
- Oyston JW, Hughes M, Gerber S, Wills MA. 2016 Why should we investigate the morphological disparity of plant clades? *Ann. Bot.* **117**, 859–879. (doi:10.1093/aob/mcv135)
- Stebbins GL. 1951 Natural selection and the differentiation of angiosperm families. *Evolution* **5**, 299–324. (doi:10.2307/2405676)
- Chartier M, Jabbour F, Gerber S, Mitteroecker P, Sauquet H, von Balthazar M, Staedler Y, Crane PR, Schönenberger J. 2014 The floral morphospace—a modern comparative approach to study angiosperm evolution. *New Phytol.* **204**, 841–853. (doi:10.1111/nph.12969)
- Lupia R. 1999 Discordant morphological disparity and taxonomic diversity during the Cretaceous angiosperm radiation: North American pollen record. *Paleobiology* **25**, 1–28.
- McGhee GR. 1999 *Theoretical morphology: the concept and its applications*. New York, NY: Columbia University Press.
- Eble G. 2000 Theoretical morphology: state of the art. *Paleobiology* **26**, 520–528. (doi:10.1666/0094-8373(2000)026<0520:TMSOTA>2.0.CO;2)
- Niklas KJ. 1999 Evolutionary walks through a land plant morphospace. *J. Exp. Bot.* **50**, 39–52. (doi:10.1093/jxb/50.330.39)
- Raup DM, Michelson A. 1965 Theoretical morphology of the coiled shell. *Science* **147**, 1294–1295. (doi:10.1126/science.147.3663.1294)
- Footo M. 1992 Paleozoic record of morphological diversity in blastozoan echinoderms. *Proc. Natl Acad. Sci. USA* **89**, 7325–7329. (doi:10.1073/pnas.89.16.7325)
- Footo M. 1994 Morphological disparity in Ordovician–Devonian crinoids and the early saturation of morphological space. *Paleobiology* **20**, 320–344. (doi:10.1017/S009483730001280X)
- Erwin DH. 2007 Disparity: morphological pattern and developmental context. *Palaeontology* **50**, 57–73. (doi:10.1111/j.1475-4983.2006.00614.x)
- Punt W, Hoen PP, Blackmore S, Nilsson, Le Thomas A. 2007 Glossary of pollen and spore terminology. *Rev. Palaeobot. Palynol.* **143**, 1–81. (doi:10.1016/j.revpalbo.2006.06.008)
- Harris JM, Hirst JL, Mossinghoff MJ. 2000. *Combinatorics and graph theory*. New York, NY: Springer.
- Glatz E, Mavromatis S, Ager B, Dimitropoulos X. 2014 Visualizing big network traffic data using frequent pattern mining and hypergraphs. *Computing* **96**, 27–38. (doi:10.1007/s00607-013-0282-8)
- Tricaud S, Nance K, Saadé P. 2011 Visualizing network activity using parallel coordinates. In *Proc. Hawaii Int. Conf. on System Sciences*, **44**, pp. 1–8. Washington, DC: IEEE Computer Society.
- Ter Steege H *et al.* 2013 Hyperdominance in the Amazonian tree flora. *Science* **342**, 1243092. (doi:10.1126/science.1243092)
- Hubbell SP, Foster RB, O’Brien ST, Harms KE, Condit R, Wechsler B, Wright SJ, Loo de Lao S. 1999 Light-gap disturbances, recruitment limitation, and tree diversity in a Neotropical forest. *Science* **283**, 554–557. (doi:10.1126/science.283.5401.554)
- Mander L, Li M, Mio W, Fowlkes CC, Punyasena SW. 2013 Classification of grass pollen through the quantitative analysis of surface ornamentation and texture. *Proc. R. Soc. B* **280**, 20131905. (doi:10.1098/rspb.2013.1905)
- Roubik DW, Moreno JE. 1991 *Pollen and spores of Barro Colorado Island*. St Louis, MO: Missouri Botanical Gardens.

21. Harley MM, Zavada MS. 2000 Pollen of the monocotyledons: selecting characters for cladistic analysis. In *Monocots: systematics and evolution* (eds KL Wilson, DA Morrison), pp. 194–213. Melbourne, Australia: CSIRO.
22. Chaloner WG. 2013 Three palynological puzzles. *Int. J. Plant Sci.* **174**, 602–607. (doi:10.1086/668225)
23. Mander L, Punyasena SW. 2014 On the taxonomic resolution of pollen and spore records of Earth's vegetation. *Int. J. Plant Sci.* **175**, 931–945. (doi:10.1086/677680)
24. Willis JC, Yule GU. 1922 Some statistics of evolution and geographical distribution in plants and animals, and their significance. *Nature* **109**, 177–179. (doi:10.1038/109177a0)
25. Burlando B. 1990 The fractal dimension of taxonomic systems. *J. Theor. Biol.* **146**, 99–114. (doi:10.1016/S0022-5193(05)80046-3)
26. Foote M. 2012 Evolutionary dynamics of taxonomic structure. *Biol. Lett.* **8**, 135–138. (doi:10.1098/rsbl.2011.0521)
27. Thompson DW. 1942 *On growth and form*, 2nd edn. London, UK: Cambridge University Press.
28. Turing AM. 1952 The chemical basis of morphogenesis. *Phil. Trans. R. Soc. Lond. B* **237**, 37–72. (doi:10.1098/rstb.1952.0012)
29. Wright S. 1932 The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In 6th *Int. Congr. Genet.*, 1, pp. 356–366. Bethesda, MD: Genetics Society of America.
30. Blackmore S, Ferguson IK. (eds) 1986 *Pollen and spores: form and function*. London, UK: Academic Press.
31. Katifori E, Alben S, Cerda E, Nelson DR, Dumais J. 2010 Foldable structures and the natural design of pollen grains. *Proc. Natl Acad. Sci. USA* **107**, 7635–7639. (doi:10.1073/pnas.0911223107)
32. Hemsley AR, Collinson ME, Kovach WL, Vincent B, Williams T. 1994 The role of self-assembly in biological systems: evidence from iridescent colloidal sporopollenin in *Selaginella* megaspore walls. *Phil. Trans. R. Soc. Lond. B* **345**, 163–173. (doi:10.1098/rstb.1994.0095)
33. Gould SJ. 1989 *Wonderful life*. New York, NY: Norton.
34. Thomas RDK, Reif W-E. 1993 The skeleton space: a finite set of organic designs. *Evolution* **47**, 341–360. (doi:10.2307/2410056)
35. Lewontin RC. 2003 Four complications in understanding the evolutionary process. *Santa Fe Inst. Bull.* **18** (centre insert, unpaginated).
36. Friis EM, Crane PR, Pedersen KR. 2011 *Early flowers and angiosperm evolution*. New York, NY: Cambridge University Press.