A combinatorial approach to angiosperm pollen morphology

A combinatorial approach to angiosperm pollen morphology

Luke Mander[1]

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

Author for correspondence:
Luke Mander
e-mail: luke.mander@open.ac.uk

Abstract

Angiosperms (flowering plants) are strikingly diverse. This is clearly expressed in the morphology of their pollen grains, which are characterised 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 visualise using a parallel coordinates plot. This represents a raw morphospace. In order 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.

Keywords:

plants, morphology, morphospace, combinatorics, pollen

1. Introduction

Angiosperms (flowering plants) are an extremely diverse group of terrestrial plants and are composed of an estimated 260,000 species [1]. Angiosperms are characterised 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 (~150–65 million years ago) [5], and recent work has highlighted that plant clades, including the angiosperms, typically attain high level of morphological disparity early in their evolutionary history [2].

In general, researchers recognise 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 1,008 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 employed 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 morphotypes reflect different combinations of these character states. This is shown in Figure 1*a* that 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]).

2.1. Counting the number of possible angiosperm pollen morphotypes

One possible combination of the character states shown in Figure 1*a* 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 the tectum that forms a roof over its internal structures [13]. This pollen grain has a spherical 3-dimensional shape, is dispersed at maturity by its parent plant as a single grain (a monad), and has spines longer than 1μm 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 posses 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

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

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 $\binom{n}{1}$ (Fig. 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 $\binom{5}{2}$ (Fig 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.

2.2. 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. In order 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 Supplementary Material. This algorithm takes lists of morphological characters as it's 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 Supplementary Material.

I have visualised these character combinations using a parallel coordinates plot (Fig. 1*c*). In this plot, each vertical axis represents a different morphological character and lines visualise 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 visualise large volumes of network traffic data [16]. This plot represents a raw morphospace that is bounded and contains a range of possible morphologies (Fig 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 Earth, and supports an estimated ~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 focussed on the morphology of angiosperm pollen grains produced by plants growing on Barro Colorado Island (BCI). Barro Colorado Island is a 1,560-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 present 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].

3.1. Morphospace construction

I used the methods outlined in Figure 1 to construct a raw angiosperm pollen morphospace based on 60 aperture character states, four exine character states, ten shape character states, six dispersal character states and 23 surface ornamentation character states (a full list of character states is provided in the 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

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

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 visualised. 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.

3.2. Patterns of morphospace occupation by real-world Neotropical angiosperm pollen grains

In order 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: (1) those

242 lacking a plate, (2) those with missing characters, and (3) polyads. I followed the
243 higher taxonomic classifications in [20], and this resulted in a dataset consisting
244 of 115 angiosperm plant families, which contain a total of 1,008 species and
245 produce a total of 468 pollen morphotypes (Table 1; a full list of taxa and
246 character scorings is provided in the Supplementary Material). This is just 0.01%
247 of the 3,643,200 pollen morphotypes that were generated algorithmically by *n*
248 choose *k* combination and used to construct the raw morphospace shown in
249 Figure 2.

250 I split these 115 families into monocots (monocotyledons; angiosperms
251 whose seeds contain a single cotyledon) and dicots (dicotyledons; angiosperms
252 whose seeds contain two cotyledons). This partitioned dataset consists of 19
253 monocot families, which contain a total of 184 species and produce 68 pollen
254 morphotypes, and 96 dicot families, which contain a total of 824 species and
255 produce 407 pollen morphotypes (Table 1). I then mapped each BCI monocot
256 and dicot pollen morphotype onto the raw morphospace shown in Figure 2. This
257 shows how the morphospace occupied by the pollen of these two plant groups
258 compares to the morphospace created by 3,643,200 algorithmically generated
259 pollen morphotypes (Fig. 3).

260 The number of real-world BCI morphotypes in this analysis is clearly too
261 few to occupy the entire raw morphospace. However, visual inspection of Figure
262 3 highlights that despite this, a considerable amount of the available
263 morphospace is occupied by the BCI pollen morphotypes (Fig. 3). This is
264 particularly noticeable in the morphospace occupied by dicot pollen, and for this
265 plant group there are some associations between character states that are
266 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 (Fig. 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 (Fig. 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 realised in the BCI dataset of monocot pollen (Fig. 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 Fig. 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 (Fig. 3). This association represents pollen grains that have just a single type of surface ornamentation. Associations between character states on the 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 (Fig. 3).

This mapping also shows that the pollen grains produced by BCI monocots and dicots occupy strikingly different regions of morphospace (Fig. 3). This is highlighted by the differences in monocot and dicot aperture character

292 states. Apertures are a key character in the study of angiosperm pollen, and are
293 thought to be a key difference between the pollen of the monocot and dicot
294 clades [21]. The aperture character states of BCI monocots are situated in two
295 groups that are positioned at opposite poles of the aperture character axis (Fig.
296 3*a*). The group at the lower extremity of the character axis contains inaperturate,
297 monoporate, triporate, pantoporate and stephanoporate morphotypes, while the
298 group at the upper extremity contains zonorate, trichotomosulcate, disulcate and
299 monosulcate morphotypes (see [13] for terminology). These two groups are
300 separated by a region of morphospace that is largely unoccupied by monocot
301 morphotypes (Fig. 3*a*). This region contains tricolpate, stephanocolpate and
302 tricolporate forms (see [13] for terminology), and is occupied by a variety of
303 dicot pollen morphotypes (Fig. 3*b*). These aperture types are highly
304 characteristic of the eudicot plant clade, but the selective pressure (if any)
305 driving their adoption by this plant group is largely unknown [22].
306
307 **4. Discussion**
308
309 **4.1. Pollen grains and Neotropical plant diversity**
310 There are fewer pollen morphotypes than plant species in the dataset of BCI
311 angiosperms that I have used in this paper. This is reflected in the number of
312 pollen morphotypes expressed as a proportion of species for all angiosperms in
313 the dataset as well as monocots and dicots (Table 1). Some of the plant families
314 that constitute the BCI flora produce very few pollen morphotypes. Such families
315 include the Poaceae (grasses), which are taxonomically very rich (53 species in
316 this BCI dataset) but only produce six pollen morphotypes (Table 1). Other

317 families are also very speciose but instead produce a large number of pollen
318 morphotypes, and an example is the Papilionoideae (a sub-family of legumes),
319 which contains 60 species in this BCI dataset and produces 47 pollen
320 morphotypes (Table 1).

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

331 Some data points are indeed situated well below the line of equality, and
332 pollen morphotypes clearly underestimate the species diversity of these families
333 (Fig. 4). However, there are 32 families that plot directly on the line of equality,
334 23 families that miss the line by just 1 (both excluding families represented by a
335 single species on BCI), and there are several other families that plot close to this
336 line of equality (Fig. 4). For these families, assemblages of dispersed pollen
337 grains will provide an accurate or at least reasonable reflection of their species
338 diversity (Fig. 4). Additionally, there are some plant species on BCI that produce
339 dimorphic pollen, and examples include *Cocos nucifera* (the coconut tree), which
340 produces pollen with either a monosulcate or a trichotomosulcate aperture.
341 Indeed, among monocots there are several families in which pollen aperture type

342 is not consistent within anthers, and these include the Agavaceae and Iridaceae
343 (both Asparagales), as well as the Arecaceae, which contains *Cocos nucifera* [21].
344 It is possible, therefore, that in certain circumstances data points could lie above
345 the line of equality in Figure 4, and that assemblages of dispersed pollen grains
346 could overestimate the diversity of the source vegetation. The patterns of
347 morphospace occupation by BCI angiosperm pollen grains (Fig. 3) therefore
348 reflect a situation in which some plant families, such as the Papilionoideae,
349 contribute a large quantity of morphological variety, whereas other taxa such as
350 the Poaceae contribute much less morphological variety.
351
352 **4.2. The structure and utility of a simplified palynological taxonomy**
353 In this paper, the morphology of angiosperm pollen grains has been encoded in a
354 system of discrete characters (see Fig. 2 and the Supplementary Material). This
355 system contains fewer characters than would be employed by a palynologist
356 seeking to classify pollen grains for the purpose of reconstruction vegetation
357 history, and omits certain characters that may be phylogenetically informative
358 such as the composition of the endexine and the nature of the intine [21].
359 Additionally, when observed using a microscope, the surface ornamentation of
360 pollen grains can appear to vary continuously between individuals and species in
361 a manner that is not well captured by the discrete approach that is taken here. As
362 such, both the total number of character combinations that are represented by
363 the raw morphospace shown in Figure 2, and the number of BCI pollen
364 morphotypes mapped onto this morphospace in Figure 3, represent minimum
365 diversity. The inclusion of just two size classes $\binom{2}{1}$, for example, would double

366 the total number of possible character state combinations used to construct the
367 raw morphospace shown in Figure 2:

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

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

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

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

373 This discrete character-based system of encoding pollen morphology is
374 comparable to formal taxonomic descriptions of pollen grains in the sense that
375 both involve breaking pollen morphology down into individual characters, which
376 are then described using specialist terminology (Fig. 1*a*) [13]. Further, in spite of
377 its simplicity, this system contains a relationship between the number of species
378 per pollen morphotype and the number of pollen morphotypes that is essentially
379 of the same mathematical form as the relationship between the number of
380 species per genus and the number of genera in other taxonomic systems (Fig. 5).
381 These include classifications of plants, sponges, molluscs, fishes, beetles and
382 birds [24, 25]. The correspondence between Figure 5 and similar plots from
383 other taxonomic systems [24, 25], which are often presented on log-log plots,
384 suggests that studying the structure of palynological taxonomy through
385 evolutionary time could provide data on the diversification history of plant
386 clades [see 26].

Together with data comparing plant species richness and the number of pollen morphotypes for all families at BCI (Fig. 4), Figure 5 indicates that angiosperm pollen morphospace at BCI (Fig. 3) has been filled mostly by pollen morphotypes that are unique to single species (301 instances; top-left data point in Fig. 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 (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 likely 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.

4.3. 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. Fig. 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 (Fig. 2; Table 1;

412 Supplementary Material) is biologically plausible and structurally possible.
413 There are no morphotypes with both tricolpate and monocolpate apertures, for
414 example, and there are no morphotypes that contain a character state that is not
415 known to exist in nature. However, the number of these plausible and possible
416 angiosperm pollen morphotypes is far greater than the number of angiosperm
417 species on Earth as a whole (~260,000; [1]) and in the BCI dataset I have used
418 here (1008; Table 1). This means that when real-world forms are mapped onto
419 the raw morphospace that contains these 3,643,200 morphotypes there are large
420 areas of unoccupied morphospace (Fig. 3). Although the BCI dataset used in this
421 study represents just a small fraction of angiosperm life, I suspect that a similar
422 analysis incorporating all known angiosperm species would produce a similar
423 pattern.
424 One way of interpreting patterns of morphospace occupation is to think of
425 the entire morphospace as a fitness landscape, with occupied areas representing
426 local optima and unoccupied areas representing less fit alternatives [12]. This
427 view of morphospace occupation is informed by Wright's studies of allele
428 combinations and fitness landscapes, in which he viewed a possible 10^{1000} allele
429 combinations as a rugged field with "an enormous number of widely separated
430 harmonious combinations" [29, p. 358]. Wright viewed these harmonious
431 combinations as adaptive peaks separated by maladaptive valleys, and Stebbins,
432 without directly referencing Wright's work, seems to interpret the patterns in his
433 floral morphospace in a similar fashion: "there are three types of combinations
434 which, although structurally possible, are found so rarely that they are
435 apparently poor adaptations, usually of low survival value" [3, p. 314].

436 Many attempts have been made to establish the function of the
437 morphological characters that together make up a pollen grain [e.g. 30], and
438 discussion has ranged from the role of surface ornamentation in pollination [22],
439 to the role of apertures in the desiccation, hydration and germination of pollen
440 grains [31]. Consequently, it is currently unclear which morphological features of
441 pollen grains are adaptive and which are not, and this means that the relative
442 roles of chance versus necessity in generating the morphological diversity of
443 pollen grains are largely unknown. Indeed, the wide range of plausible functional
444 roles for the morphological features of pollen grains, suggests that they are "not
445 'optimally' designed for a specific function, but merely structures that work with
446 varying efficiency in a specific ecological and evolutionary context" [30, p. 182].
447 With this in mind, I do not interpret the areas of unoccupied morphospace in
448 Figure 3 as regions containing pollen morphotypes that are maladaptive in the
449 context of the BCI flora. It is possible, however, that further experimental work
450 on the function of specific morphological features may reveal regions of
451 morphospace that are characterised by forms with specific functional properties.
452 Such work could also include analyses of the tendency of certain morphological
453 features to arise by self-assembly processes, which may produce forms that
454 reflect convergence upon a minimum free energy state rather than the effects of
455 natural selection [32].
456 Alternative interpretations of morphospace occupation include the
457 suggestion that the present-day distribution of morphology (e.g. Fig. 3) reflects
458 an earlier more continuous distribution that has been subsequently fragmented
459 by extinction [12], as well as Gould's classic idea that chance played a major role
460 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. Fig. 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 Fig. 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 to 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 (Fig S1, Fig, 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; Fig. 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.

In contrast, there are relatively few morphotypes that are repeated in three–nine families (Fig. 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–nine families share this character state (Fig. 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 (Fig. 6). These morphotypes also have a very uniform set of aperture characters: two are inaperturate and five have multiple pores (pantoporate) (Fig. 6). As quoted in

511 [22] "The systematic distribution of periporate [pantoporate of many authors]
512 pollen in extant angiosperms indicates that this aperture configuration evolved
513 independently in many different groups" [36, p. 406]. The repetition of these
514 aperture configurations across a fundamental divide in plant phylogeny is an
515 example of evolutionary convergence, and in this respect, the BCI flora reflects a
516 broader pattern in angiosperm evolution.
517
518 **5. Concluding remarks**
519 The combinatorial approach to angiosperm pollen morphology that I have
520 outlined in this paper provides an example of how the general challenge of
521 analysing large numbers of discrete character combinations [3] can be overcome.
522 This approach generates biological forms by algorithmically combining character
523 states and enumerating the results of these combinations (Fig. 1). I have used
524 this approach to generate 3,643,200 angiosperm pollen morphotypes and
525 produce a raw morphospace (Fig. 2). Some of these morphotypes correspond to
526 real-world forms whereas others do not.
527 I have mapped the pollen of 1008 species of Neotropical angiosperms
528 growing on BCI, Panama, onto this raw morphospace (Fig. 3). This allows
529 realized and unrealized forms to be compared, and shows that despite the
530 relatively small size of the BCI dataset, these real-world pollen grains occupy a
531 considerable amount of the raw morphospace (Fig. 3). This highlights that in
532 addition to their well-documented taxonomic diversity, Neotropical rainforests
533 also represent an enormous reservoir of morphological diversity. Some plant
534 families, such as the Papilionoideae, contribute a large amount of morphological

535 diversity, whereas other taxa such as the Poaceae contribute much less
536 morphological diversity (Fig. 4).
537 In this paper, the morphology of angiosperm pollen grains has been
538 encoded in a highly simplified system of five discrete characters. Despite this
539 simplicity, this system contains a relationship between the number of species per
540 pollen morphotype and the number of pollen morphotypes that has essentially
541 the same mathematical form as the relationship between the number of species
542 per genus and the number of genera in other taxonomic systems (Fig. 5). This
543 highlights that angiosperm pollen morphospace at BCI (Fig. 3) has been filled
544 mostly by pollen morphotypes that are unique to single species (Fig. 5).
545 Angiosperm pollen morphotypes that are repeated among higher taxa at BCI
546 reflect both phylogenetic constraint and evolutionary convergence (Fig. 6).
547
548 **Data accessibility.** The datasets and code supporting this article have been
549 uploaded as part of the Supplementary Material.
550
551 **Competing interests.** I declare I have no competing interests.
552
553 **Author contributions.** LM conceived, designed and undertook the study. LM
554 wrote the paper.
555
556 **Acknowledgements.** I am grateful to Claire Belcher and Carlos Jaramillo who
557 provided useful comments on an earlier version of this work. I am grateful to the
558 reviewers of this work, particularly Roger Thomas, whose thoughtful
559 suggestions clarified and improved the ideas presented here.

Funding. I received no funding for this work.

References

1. Soltis PS, Soltis DE. 2004 The origin and diversification of angiosperms. *Am. J. Bot.* **91**, 1614–1626.
2. Oyston JW, Hughes M, Gerber S, Wills MA. 2016 Why should we investigate the morphological disparity of plant clades? *Ann. Bot.* **117**, 859–879.
3. Stebbins GL. 1951 Natural selection and the differentiation of angiosperm families. *Evolution* **5**, 299–324.
4. Chartier M, Jabbour F, Gerber S, *et al*. 2014 The floral morphospace – a modern comparative approach to study angiosperm evolution. *New Phyt.* **204**, 841–853.
5. Lupia R. 1999 Discordant morphological disparity and taxonomic diversity during the Cretaceous angiosperm radiation: North American pollen record. *Paleobiology* **25**, 1–28.
6. McGhee GR. 1999 *Theoretical Morphology: the Concept and its Applications*. New York, NY: Columbia University Press.
7. Eble G. 2000 Theoretical morphology: state of the art. *Paleobiology* **26**, 520–528.
8. Niklas KJ. 1999 Evolutionary walks through a land plant morphospace. *J. Exp. Bot.* **50**, 39–52.
9. Raup DM, Michelson A. 1965 Theoretical morphology of the coiled shell. *Science* **147**, 1294–1295.

10. Foote M. 1992 Paleozoic record of morphological diversity in blastozoan echinoderms. *Proc. Natl Acad. Sci. USA* **89**, 7325–7329.
11. Foote M. 1994 Morphological disparity in Ordovcian–Devonian crinoids and the early saturation of morphological space. *Paleobiology* **20**, 320–344.
12. Erwin DH. 2007 Disparity: morphological pattern and devlopmental context. *Palaeontology* **50**, 57–73.
13. Punt W, Hoen PP, Blackmore S, Nilsson, Le Thomas A. 2007 Glossary of pollen and spore terminology. *Rev. Palaeobot. Palynol.* **143**, 1–81.
14. Harris JM, Hirst JL, Mossinghoff MJ. 2000. *Combinatorics and Graph Theory*. New York, NY: Springer-Verlag.
15. Glatz E, Mavromatidis S, Ager B, Dimitropoulos X. 2012 Visualizing Big Network Traffic Data using Frequent Pattern Mining and Hypergraphs. *Computing* **96**, 27–38.
16. Tricaud S, Nance K, Saadé P. 2011 Visualizing network activity using parallel coordinates. *Proc. Hawaii International Conference on System Sciences 44***,** 1–8.
17. ter Steege H, Pitman NCA, Sabatier D, *et al.* 2013 Hyperdominance in the Amazonian tree flora. *Science* **342**, 1243092.
18. Hubbell SP, Foster RB, O'Brien ST, Harms KE, Condit R, Wechsler SJ, Wright S, Loo de Lao S. 1999 Light-gap disturbances, recruitment limitaiton, and tree diversity in a Neotropical forest. *Science* **283**, 554–557.

608 19. Mander L, Li M, Mio W, Fowlkes CC, Punyasena SW. 2013 Classification of
609 grass pollen through the quantitative analysis of surface ornamentation
610 and texture. *Proc. R. Soc B* **280**, 20131905.
611 20. Roubik DW, Moreno JE. 1991 *Pollen and spores of Barro Colorado Island*.
612 St. Louis, Missouri: Missouri Botanical Gardens.
613 21. Harley MM, Zavada MS. 2000 Pollen of the monocotyledons: selecting
614 characters for cladistic analysis. In *Monocots: Systematics and Evolution*
615 (eds KL Wilson, DA Morrison), pp. 194–213. Melbourne, Australia: CSIRO.
616 22. Chaloner WG. 2013 Three palynological puzzles. *Int. J. Plant Sci.* **174**, 602–
617 607.
618 23. Mander L, Punyasena SW. 2014 On the taxonomic resolution of pollen and
619 spore records of Earth's vegetation. *Int. J. Plant Sci.* **175**, 931–945.
620 24. Willis JC, Yule GU. 1922 Some statistics of evolution and geographical
621 distribution in plants and animals and their significance. *Nature* **109**,
622 117–179.
623 25. Burlando B. 1990 The fractal dimension of taxonomic systems. *J. Theor.*
624 *Biol.* **146**, 99–114.
625 26. Foote M. 2012 Evolutionary dynamics of taxonomic structure. *Biol. Lett.* **8**,
626 135–138.
627 27. Thompson DW. 1942 *On Growth and Form*. 2nd Edition. London, UK:
628 Cambridge University Press.
629 28. Turing AM. 1952 The chemical basis of morphogenesis. *Phil. Trans. R. Soc.*
630 *Lond. B* **237**, 37–72.
631 29. Wright S. 1932 The roles of mutation, inbreeding, crossbreeding, and
632 selection in evolution. *Int. Congr. Genet. 6* **1**, 356–366.

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.

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.

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.

35. Lewontin RC. 2003 Four complications in understanding the evolutionary process. *Santa Fe Institute Bulletin* **18**, (centre insert, unpaginated).

36. Friis EM, Crane PR, Pedersen KR. 2011 *Early Flowers and Angiosperm Evolution*. New York, NY: Cambridge University Press.

Figure and table captions

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*) Showing how these different combinations can be represented using a parallel coordinates plot.

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 Supplementary Material.

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 Supplementary Material.

Figure 4. Scatterplot comparing the species richness of angiosperm families from Barro Colorado Island with the number of pollen morphotypes produced by each of these families. Datapoints have been plotted for each of the 19 monocot and 96 dicot families analysed in this paper but datapoints for certain families overlap. Both axes are plotted on a logarithmic scale. Diagonal dashed line represents a line of equality.

683 **Figure 5.** Scatterplot showing the hollow-curve relationship between the
684 number of species per angiosperm pollen morphotype and the number of
685 angiosperm pollen morphotypes on Barro Colorado Island.
686
687 **Figure 6.** Proportions of angiosperm pollen morphotypes that are repeated
688 among higher taxa at BCI (green vertical bars). Numbers enclosed within grey
689 vertical boxes show the distribution of aperture character states that are
690 repeated among higher taxa at BCI (see dataset S1 for a full list of aperture
691 character states). Key aperture character states repeated among three–nine
692 families as follows: Inaperturate (1), Pantoporate (11), Tricolpate (25),
693 Tricolporate (44), Tricolporate Marginate (45), Pantocolporate (48),
694 Monosulcate (54). Cartoons to the right are schematic pollen grains that
695 graphically display each of these aperture character states.
696
697 **Table 1.** Summary of the data and results discussed in this paper.
698
699 **Supplementary material**
700 **Dataset S1.** List of characters used to generate the raw morphospace shown in
701 Figure 2.
702
703 **Dataset S2.** Taxa and character scorings of BCI pollen grains used to generate
704 Figure 3. Refer to Dataset S1 for list of characters.
705
706 **Supplementary File S1.** An example Python algorithm that counts and
707 enumerates all possible combinations of the character states shown in Figure 1*a*.

708

709 **Supplementary File S2.** An example of a .csv file containing all possible

710 combinations of the character states (in numerical form) shown in Figure 1*a*.

711

712 **Figure S1.** Barchart of the frequency of angiosperm pollen morphotypes at BCI.

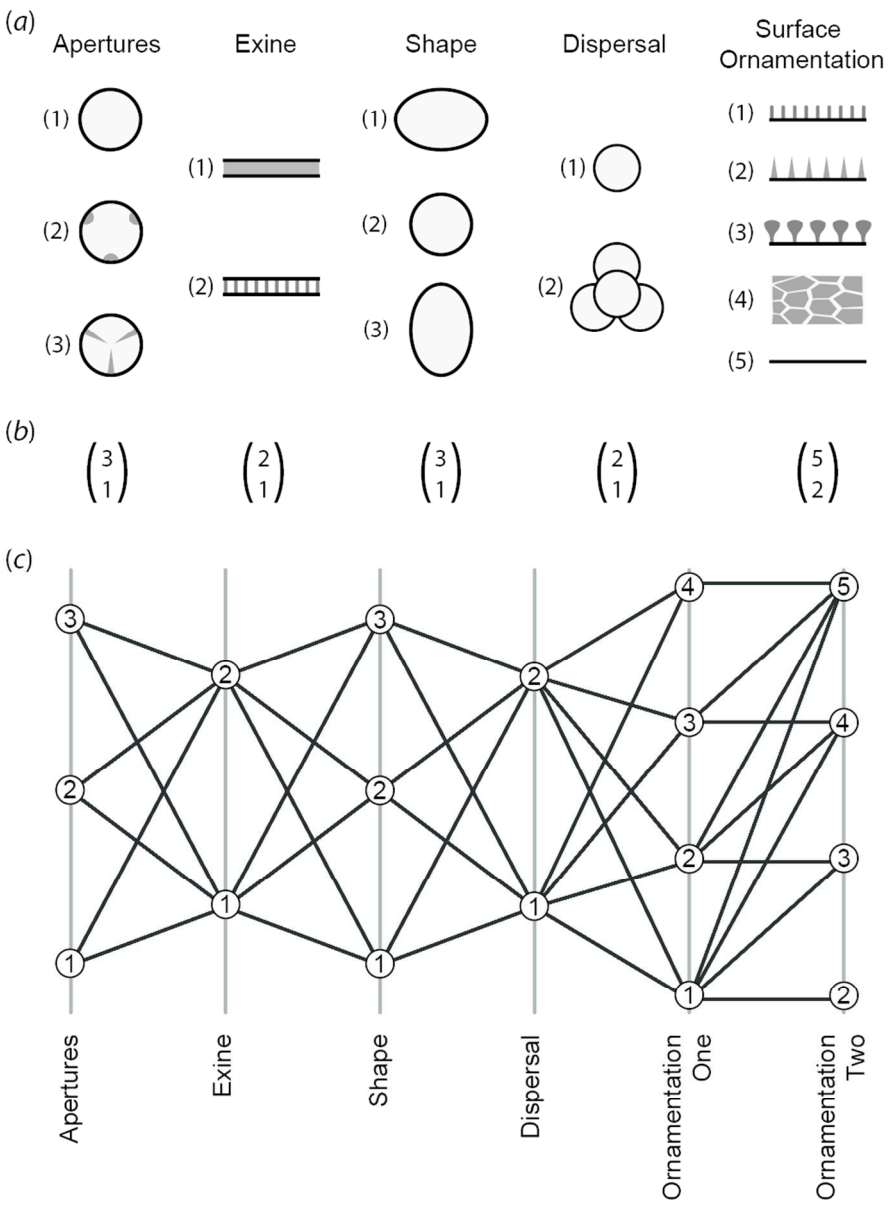

Figure 1. (a) Schematic illustration of characters that make up an angiosperm pollen grain. Aperture character states: inaperturate (1), triporate (2), tricolpate (3). Exine character states: atectate (1), tectate (2). Shape character states: oblate (1), spheroidal (2), prolate (3). Dispersal character 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) Showing how these different combinations can be represented using a parallel coordinates plot.

88x118mm (300 x 300 DPI)

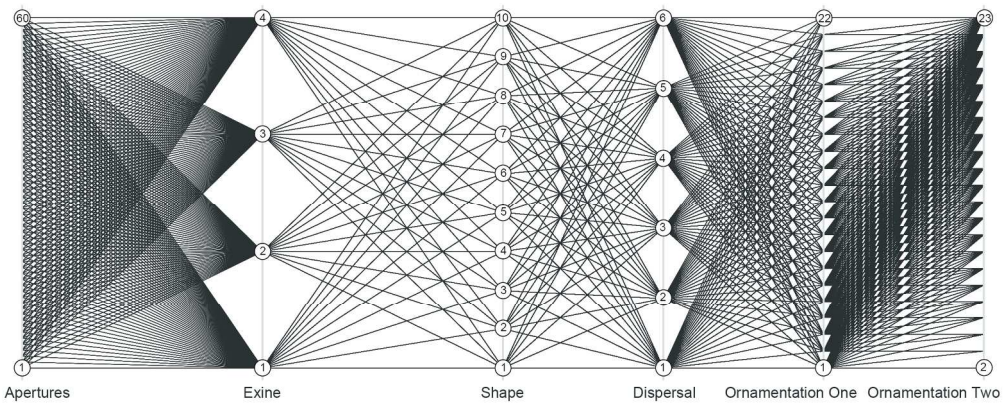

Figure 2. Parallel coordinates plot of 3,643,200 angiosperm pollen morphotypes generated algorithmically by n choose k combination. Numbers enclosed within white circles represent each of the character states used to produce this morphospace. Only the first and last aperture and surface ornamentation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

187x75mm (300 x 300 DPI)

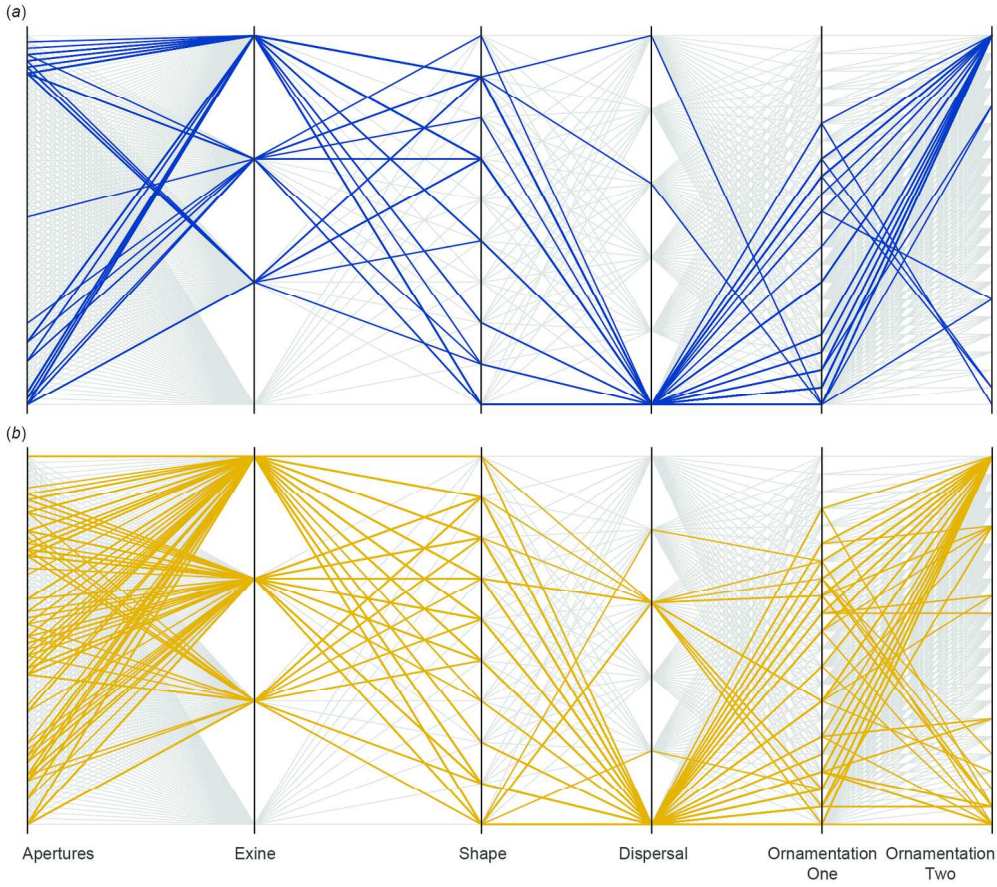

Figure 3. Parallel coordinates plots showing how pollen grains produced by angiosperms growing on Barro Colorado Island map onto the morphospace shown in Figure 2. In both plots this morphospace is shown in light grey, and is overlain by (a) monocot pollen morphotypes in dark blue and (b) dicot pollen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

170x151mm (300 x 300 DPI)

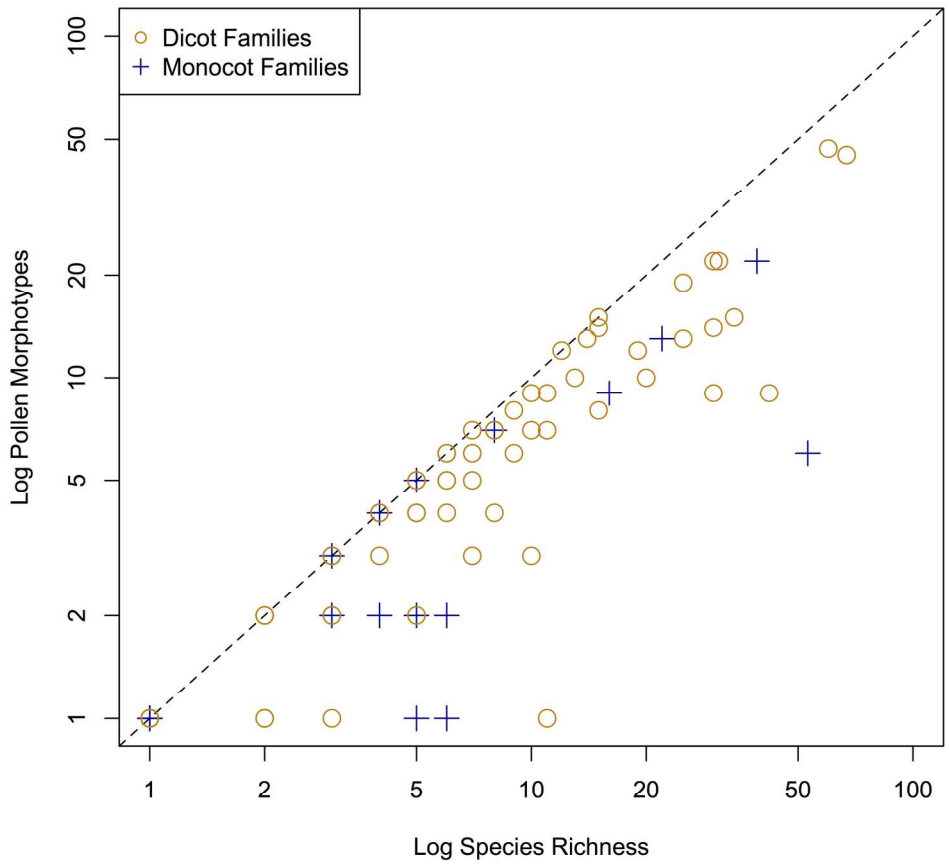

Figure 4. Scatterplot comparing the species richness of angiosperm families from Barro Colorado Island with the number of pollen morphotypes produced by each of these families. Datapoints have been plotted for each of the 19 monocot and 96 dicot families analysed in this paper but datapoints for certain families overlap. Both axes are plotted on a logarithmic scale. Diagonal dashed line represents a line of equality.

177x177mm (300 x 300 DPI)

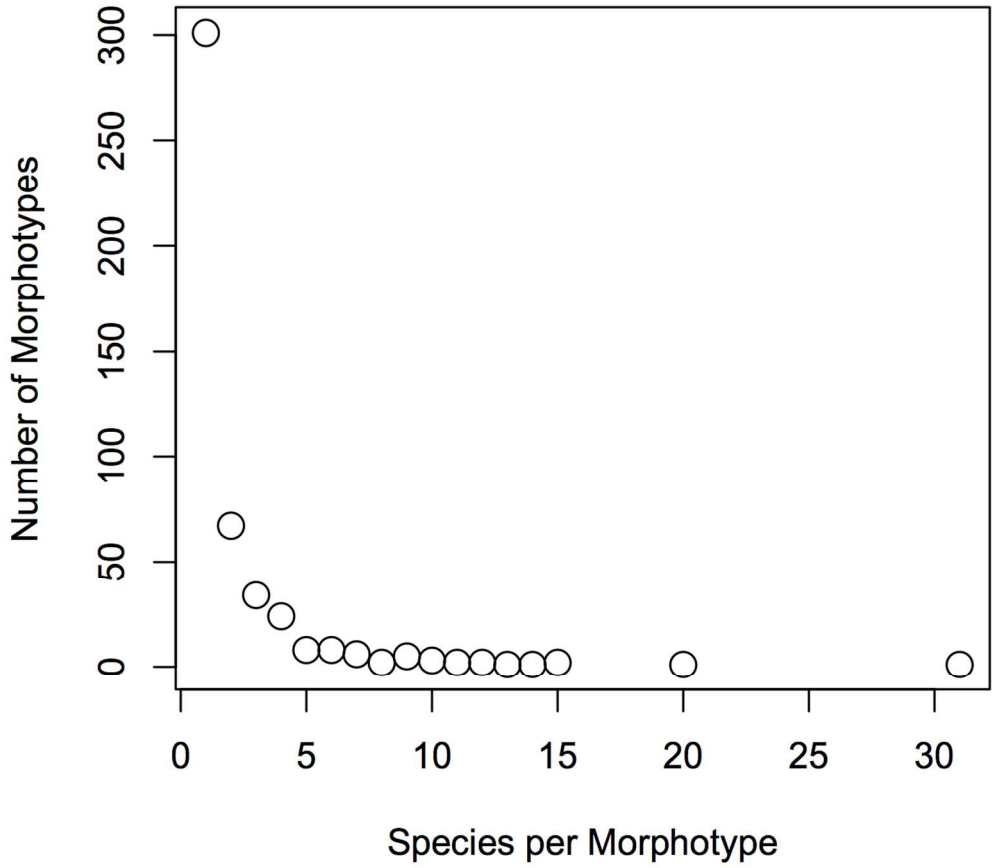

Figure 5. Scatterplot showing the hollow-curve relationship between the number of species per angiosperm pollen morphotype and the number of angiosperm pollen morphotypes on Barro Colorado Island.

117x103mm (300 x 300 DPI)

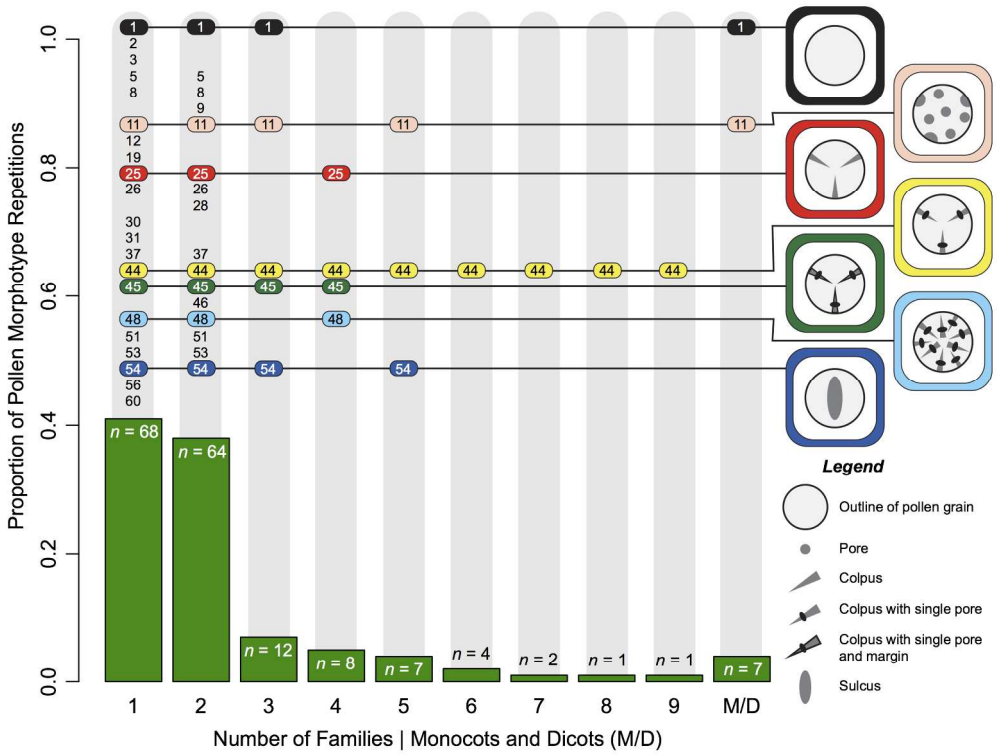

Figure 6. Proportions of angiosperm pollen morphotypes that are repeated among higher taxa at BCI (green vertical bars). Numbers enclosed within grey vertical boxes show the distribution of aperture character states that are repeated among higher taxa at BCI (see dataset S1 for a full list of aperture character states). Key aperture character states repeated among three–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.

204x153mm (300 x 300 DPI)

	Families (*n*)	Species (*n*)	Morphotypes (*n*)	Morphotypes as a proportion of Species
n choose *k* 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