Electronic Supplementary Material

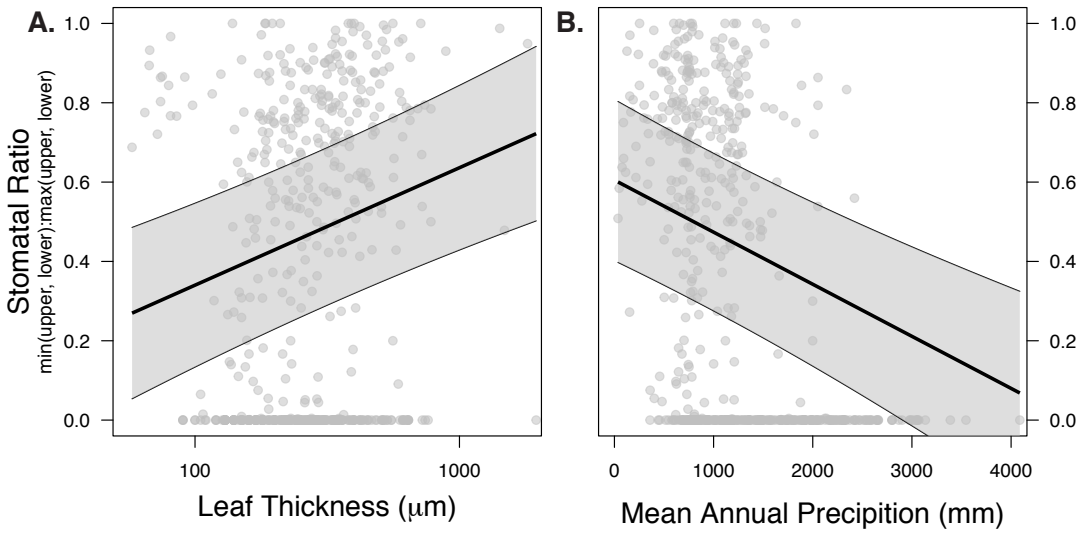

Fig. S1. Amphistomy is weakly associated with thicker leaves and drier habitats. Each point represents a species from the global dataset. The thick line and gray polygon are the median and 95% confidence intervals from the posterior distribution of predicted stomatal ratio as a function of leaf thickness or mean annual precipitation based on phylogenetic regression. The fitted lines and confidence intervals are drawn with growth form set to 'perennial' and other continuous predictor variables set to their median.

Table S1. Multiple selective regimes are manifest in a multimodal trait distribution. Models with multiple components (k) corresponding to distinct selective regimes under a bounded Ornstein-Uhlenbeck process fit the data significantly better than models with a single regime (lower Bayesian Information Criterion [BIC]). In particular, the model with with three regimes is much more strongly supported than models with one or two regimes (see Fig. 1B–D for a visual representation of regimes). A mixture of multiple regimes, in turn, gives rise to a multimodal distribution with hypo-, amphi-, and hyperstomatous modes. For a given mixture, each of k regimes is represented as a component i parameterized by the strength of constraint (ϕ_i) around the long-term average (θ_i) and a mixture weight w_i.

k	Parameters			log-likelihood	df	BIC
1	$\phi_1 = 0.4$	$\theta_1 = 0.17$	$w_1 = 1$	-604	2	1220.9
2	$\phi_1 = 0.25$	$\theta_1 = 0.04$	$w_1 = 0.52$	-252.5	5	536.9
	$\phi_2 = 9.98$	$\theta_2 = 0.46$	$w_2 = 0.48$			
3	$\phi_1 = 0.16$	$\theta_1 = 0.02$	$w_1 = 0.47$	-237.7	8	526.6
	$\phi_2 = 17.24$	$\theta_2 = 0.47$	$w_2 = 0.38$			
	$\phi_3 = 2.04$	$\theta_3 = 0.35$	$w_3 = 0.16$			
4	$\phi_1 = 6.99$	$\theta_1 = 0$	$w_1 = 0.44$	-235.6	11	541.6
	$\phi_2 = 1.6$	$\theta_2 = 0.35$	$w_2 = 0.17$			
	$\phi_3 = 16.85$	$\theta_3 = 0.47$	$w_3 = 0.38$			
	$\phi_4 = 181.8$	$\theta_4 = 0.99$	$w_4 = 0$			

Table S2. Growth form, anatomy, and precipitation jointly determine stomatal ratio. Three models with varying levels of phylogenetic signal (Brownian motion [top], Pagel's λ [middle], and nonphylogenetic [bottom]) identify growth form, leaf thickness, and mean annual precipitation as significantly associated with stomatal ratio.

Stomatal Ratio $\sim$	df	SS	MS	F	P
Brownian Motion					
log(Leaf Thickness)	1	0.017	0.017	20.31	8.08×10^{-6}
Mean Annual Precipitation	1	0.021	0.021	24.11	1.21×10^{-6}
Elevation	1	0	0	0.08	0.78
Leaf Area Index	1	0	0	0.05	0.82
Growth Form	5	0.039	0.008	9.06	2.74×10^{-8}
Pagel's $\lambda = 0.64$					
log(Leaf Thickness)	1	0.008	0.008	24.38	1.05×10^{-6}
Mean Annual Precipitation	1	0.009	0.009	26.03	4.67×10^{-7}
Elevation	1	0	0	0.26	0.61
Leaf Area Index	1	0	0	0	1
Growth Form	5	0.027	0.005	15.52	2.77×10^{-14}
Nonphylogenetic					
log(Leaf Thickness)	1	2.376	2.376	31.67	2.94×10^{-8}
Mean Annual Precipitation	1	1.711	1.711	22.81	2.31×10^{-6}
Elevation	1	0.009	0.009	0.12	0.72
Leaf Area Index	1	0.031	0.031	0.41	0.52
Growth Form	5	15.897	3.179	42.38	7.36×10^{-37}

Appendix S1: Hypothesized benefits and costs of amphistomy

There are at least seven viable, non-mutually exclusive hypotheses for on the adaptive significance of amphistomy, five of which I evaluate here.

H1: Leaf thickness

The most widely cited and frequently tested diffusional limitation hypothesis is that amphistomy is adaptive in thick leaves. Models (1; 2) and experiments (3) demonstrate that the path length from substomatal cavities to chloroplasts can impose a large constraint on photosynthesis, especially when leaf thickness exceeds approximately 300 μm. Several studies have found a positive correlation between leaf thickness and amphistomy (1; 4; 5; 6; 7; 8; 9), but the evidence is equivocal (10; 11; 12).

H2: Light

A second hypothesis is that amphistomy is favoured in high light, open environments because CO_2 becomes more limiting at high irradiance. H1 and H2 are difficult to disentangle, and could even reinforce one another, because leaf thickness increases under high irradiance (13). However, several studies have argued that the light environment, rather than leaf thickness, is the primary factor affecting selection on amphistomy (14; 15; 16; 11; 4; 17).

H3: Precipitation

Wood (18) observed that amphistomy was common in Australian deserts. Although amphistomy is sometimes common in dry environments, most studies conclude that precipitation is indirectly correlated with amphistomy because drier habitats also tend to be more open (14; 17). Nevertheless, the fact that amphistomy can increase water-use efficiency (1; 19) suggests that it might be favoured in dry habitats, independent of other factors.

H4: Altitude

Anatomical surveys demonstrate that amphistomy is sometimes more common in high elevation communities compared to nearby low elevation communities (20; 21; 22), possibly because lower CO_2 partial pressures place a greater premium on efficient diffusion. However, this hypothesis is complicated by the fact that diffusion coefficients are higher at elevation because the air is thinner (23), meaning that CO_2 diffusion could actually be less limiting.

H5: Growth form

Independent of leaf anatomy and the abiotic environment, the strength of selection on photosynthetic rate might be stronger among certain growth forms (e.g. forbs vs. trees) because of their different life history strategies. Salisbury (1927) noted qualitatively that herbs tended to amphistomatous, an observation later confirmed by Peat and Fitter (1994). However, other reviews have argued that stomatal ratio

is not closely connected with any particular growth form (24; 11).

Two hypotheses I have not considered because of methodological limitations are that amphistomy is associated with vertically-oriented, isobilateral leaves (24) and that amphistomy, by doubling the conductive leaf surface area, relieves a constraint the stomatal size-density tradeoff (25; 8). I did not have sufficient, reliable information on leaf orientation and guard cell size to evaluate these hypotheses.

Costs of upper stomata

This study reaffirms at a global scale that most species are hypostomatous. The most parsimonious explanation for the preponderance of hypostomy is that there is cost to having stomata on the upper surface of the leaf. A fitness cost associated with increased evaporation (26) cannot explain the dearth of stomata on the upper leaf surface, though this explanation occasionally appears in the literature (27). In fact, amphistomy is common in some dry habitats (18; 1; 14; 4) and amphistomatous plants can be functionally hypostomatous when stressed by regulating stomatal aperture differentially on each surface (28; 29; 30; 19). Although amphistomatous plants can be functionally hypostomatous, the reverse is not true. Hence, anatomical amphistomy should be favoured whenever the capacity to be functionally amphistomatous is advantageous.

Besides evaporation, several fitness costs have been suggested, including decreased water-use efficiency of amphistomy in large leaves (1), photodamage to guard cell chloroplasts (W.K. Smith, pers. comm.), occlusion of upper stomata by water blockage (31), and increased susceptibility to foliar pathogens (2). Increased evaporation

is an unlikely explanation since so many desert species are anatomically amphistomatous (see above), but to my knowledge, most other hypotheses have not been rigorously tested. However, (32) showed that adaxial (upper) stomata pore area, but not abaxial (lower) pore area, was strongly correlated with susceptibility to a rust pathogen. Hence, the pathogen susceptibility hypothesis is best supported by the current data.

Appendix S2

Assembling a comparative data set

Stomatal ratio and leaf thickness I collected quantitative data on stomatal ratio and leaf thickness from previously published studies (see Appendix S6 for full list of data sources). These data are spread across a large and diverse literature, including functional ecology, taxonomy, agriculture, and physiology. Hence, neither a standardized nor exhaustive search was possible. I started by using Web of Knowledge to locate studies that cited seminal papers on the adaptive significance of amphistomy, specifically (1) and (11). Once I found a paper with data, I examined papers that cited those ones. Finally, I found additional data sources in comprehensive reviews of plant anatomy (33; 24; 34). For all data papers, I recorded the mean leaf thickness, abaxial (lower) and adaxial (upper) stomatal density for each species. Where only ranges were given, I used the midpoint. If the study included a treatment, I collected only data from the control treatment. If studies measured both juvenile and adult leaves, I used only adult leaves (no study reported only juvenile leaves). Usually data were given in a table, but occasionally I used ImageJ (35) to extract data from figures or contacted authors for data. I only included data from studies that intentionally examined both surfaces for stomata; I excluded data from studies that described species categorically as "hypostomatous", or "amphistomatous", or "hyperstomatous". Excluding qualitative data was necessary because there is no standard definition of "amphistomy" – it has sometimes been used to describe species that have approximately equal densities on each side (1) and at other times for species that have any stomata on the both surfaces (36; 37).

Climate and elevation Based on the *a priori* hypotheses, I extracted data on mean annual precipitation (average 1950 – 2000), elevation (Worldclim (38)), and light environment (average leaf area index between 1982 – 1998 based on remote sensing (39)). For light environment, I used a satellite indicator of leaf area index, the number of leaf layers between the ground and top of the canopy. Lower leaf area index is interpreted as a more open light environment. The strength of these global data sources is that I was able to obtain data for every species from the same dataset. A limitation of these data is that even the highest resolution ($\approx$ 1 km) data might miss important temporal and microsite variation. I discuss these limitations in light of the findings in the Discussion. For climate and elevation, geographic coordinates for each species are needed. For this, I downloaded all georeferenced herbarium specimens for a given species from GBIF (last accessed Jan 15–18, 2015) using the occ_search function in R package **rgbif** (40). I filtered out or manually edited clearly erroneous locations (e.g. lat = 0 or lon = 0 or where lat and lon were clearly reversed). The mean and median number of GBIF georeferenced occurrences per species was 737 and 194, respectively. I calculated the trimmed-mean (10% trim) mean annual precipitation, elevation, and leaf area index to further remove specimens well outside the species' range, possibly because they were, say, misidentified, cultivated, or improperly georeferenced.

Growth Form I partitioned species by growth form into the following categories: trees, small trees/shrubs, shrubs, and herbaceous species (forbs and grasses). Herbaceous species were further subdivided into annuals, biennials, and perennials. Species that were variable or intermediate (e.g. annual/biennial, annual/perennial, biennial/perennial, or annual/biennial/perennial) were classified as 'biennial'. Subshrubs

with some woody growth were lumped with perennials rather than shrubs. Where possible, I obtained growth form data from associated data papers. When this information was not given, I used regional floras, supplemented by online trait databases such as USDA Plants (41) and Encyclopedia of Life (42). When these sources were unavailable or ambiguous for a given species, I checked the primary taxonomic literature by searching the species name in Google Scholar.

Taxonomic name resolution I submitted taxonomic names in the database to the Taxonomic Name Resolution Service (TNRS) (43). I used names given by TNRS when it returned an accepted name or synonym with overall score greater than 0.97 (scores are between 0 to 1). I scrutinized names where TNRS deemed the name illegitimate, gave no opinion, or was otherwise ambiguous. At that point, I consulted additional plant taxonomic repositories: The Plant List (44), International Plant Names Index (45), and the Euro+Med PlantBase (46). When no accepted names were identified, I used original name given by the authors. For two very recent papers with up to date taxonomy by experts (8; 47), I used the names given by those authors.

Testing adaptive hypotheses for stomatal ratio using phylogenetic regression

For this analysis I quantified stomatal ratio as min(upper density, lower density):max(upper density, lower density). In this form, stomatal ratio equals 1 when the densities on each surface are the same, and goes to 0 as the distribution become more asymmetrical (hypostomy or hyperstomy). Note that this form differs from what I use in

analyzing multimodality because I wanted to specifically test which factors favour the phososynthetically optimal distribution (amphistomy) versus suboptimal distributions (either hypo- or hyperstomy). I constructed a phylogeny for species in the dataset using a Phylomatic (48) megatree approach. To examine whether results were robust to phylogenetic correction, I analyzed the data using three methods: Brownian motion (high phylogenetic signal), Pagel's λ (intermediate phylogenetic signal), and no phylogenetic signal (normal ANOVA). For the intermediate signal model, I estimated Pagel's λ using maximum likelihood. Phylogenetic models were fit using phylogenetic least squares in the R package **caper** (49).The trait dataset and phylogeny used in these analyses are available on Dryad (50).

Appendix S3: An evolutionary process model for proportion traits

Making evolutionary sense of a biological pattern requires an underlying process model to provide the theoretical foundation on which data analysis rests. A powerful approach in macroevolution involves modelling trait evolution on phenotypic landscapes with or without constraint (51; 52; 53). If models with constraint describe the data better than those without, then there is compelling evidence that phenotypic landscapes are shaped by some combination of selective, genetic, functional, or developmental constraints. Furthermore, phenotypic landscapes may change under multiple regimes, meaning that a trait is best described by a mixture of distributions, each generated under separate regimes (54; 55; 56). Current evolutionary process models such as Brownian motion and Ornstein-Uhlenbeck assume that traits follow a Gaussian distribution, but this is clearly inappropriate for traits like stomatal ratio. In this section, I modify previous evolutionary process models to accommodate proportion traits and derive the expected pattern given phenotypic landscapes that are constrained versus those that are unconstrained. This model provides a strong theoretical foundation for the model-based statistical inference described below. A glossary of symbols used in this text are provided in Table S3.

In both models with and without constraint, I assume that *total* stomatal density follows a random walk over macroevolutionary time, though the exact process is irrelevant here. Imagine for a set area (A_{leaf}) of leaf (e.g. 1 μm^2) there are $N_T(t) = A_{\text{leaf}}D_T(t) = A_{\text{leaf}}(D_U(t) + D_L(t))$, where $N_T(t)$ is the total number of stomata in that area at time t. Total stomatal number $N_T(t)$ is the sum of upper ($N_U(t)$)

Table S3. Glossary of symbols used in process models of stomatal trait evolution.

Symbol	Description
r	Stomatal ratio: ratio of upper to total stomatal density
N_T, N_U, N_L	Number of stomata in a focal leaf area A_L
	The total number N_T is the sum of upper N_U and lower N_L stomata
D_T, D_U, D_L	Density of stomata in total, upper, and lower surfaces
A_{leaf}	Focal leaf area
ν	Diffusion coefficient of stomatal ratio
θ	Long-run average stomatal ratio
α	Return rate to long-run average ratio
ϕ	Defined as $\nu\alpha$
$M_{\delta x}$	Drift function of stomatal ratio r in diffusion approximation
$V_{\delta x}$	Diffusion function of stomatal ratio r in diffusion approximation

and lower ($N_L(t)$) stomata. Let $\Delta N_{T,t} = N_T(t+1) - N_T(t)$ be the change in total stomatal number that must be made up of changes in upper stomata, lower stomata, or some combination of both. I assume that the contribution to $\Delta N_{T,t}$ from upper and lower stomata is proportional to their density. For reasons explained below, I define $\nu = N_T(t+1)$ as the total stomata at time $t+1$. The transition rate u_{ij} from $N_U = i$ upper stomata at time t to $N_U = j$ upper stomata at time $t+1$ is binomially distributed with a rate determined by the stomatal ratio r:

$$u_{ij} = \binom{\nu}{j} r^j (1-r)^{\nu-j} \quad j \in \{0, 1, 2, \ldots, \nu\} \tag{S1}$$

Note that stomatal ratio here is defined as the proportion of upper stomata, $r = N_U/(N_U + N_L) = N_U/N_T = N_U/\nu$. What this assumption says is that increasing stomatal density (upper or lower) from 100 to 120 is much easier than increasing density from, say, 0 to 20 or 10 to 30. Formally, the mean and variance of stomatal

ratio in the next time step is therefore:

$$\mu(r) = \mathbb{E}\left[\frac{N_U}{\nu}\right] = r \tag{S2}$$

$$\sigma^2(r) = \mathbb{E}\left[\left(\frac{N_U}{\nu}\right)^2\right] - \left(\mathbb{E}\left[\frac{N_U}{\nu}\right]\right)^2 = \frac{r(1-r)}{\nu} \tag{S3}$$

In other words, the average stomatal ratio does not change, but the variance increases each time step. When ν is large, the distribution can be approximated with a normal distribution and a diffusion approximation can be used to model the long term evolution of the trait. This diffusion process is analogous to Brownian motion, except that the trait is bounded by 0 and 1. It is also mathematically equivalent to one-locus, two-allele population genetic models of neutral evolution (see (57) for a detailed derivation). I will make reference to results from this literature without rigorously deriving them here. In particular, it has been shown that the stationary distribution of the diffusion is:

$$f(r) = \frac{e^{A(r)}\left(c_1 \int e^{-A(r)} \mathrm{d}r + c_2\right)}{V_{\delta x}} \tag{S4}$$

where

$$A(r) = \int \frac{2M_{\delta x}}{V_{\delta x}} \mathrm{d}r \tag{S5}$$

$$M_{\delta x} = 0 \tag{S6}$$

$$V_{\delta x} = \frac{r(1-r)}{\nu} \tag{S7}$$

and the time scale is in units of ν^{-1}. Thus, ν can be interpreted as a diffusion coefficient without necessarily specifying a genetic or developmental mechanism that governs the amount of variance in stomatal ratio from one time to the next. Solving for $f(r)$ without constraint on stomatal ratio yields:

$$f(r) = \frac{6}{r(1-r)} \tag{S8}$$

Thus, without constraint on stomatal ratio, most species should be hypo- or hyperstomatous under this model (Fig. S2), as these act like absorbing boundaries. This shows, perhaps surprisingly, that bounded traits evolving without constraint behave very differently from standard quantitative traits, which are usually expected to follow a unimodal Gaussian distribution. A plausible alternative approach to absorbing boundaries is to assume reflecting boundaries. In other words, species don't get stuck at 0 or 1, but can 'bounce' back into the middle of the distribution. In this case, the stationary distribution of the trait is uniform between 0 and 1. Thus, even with reflecting boundaries, one would not expect bounded traits to follow a distribution with an interior mode in the absence of phenotypic constraint.

Next, I modify the model to include constraint around a long-run average θ, which

may be interpreted an optimum of a selective regime. This process model is analogous to an Ornstein-Uhlenbeck process for a proportion trait. I again use the diffusion approximation, but this time the drift and diffusion coefficients are:

$$M_{\delta x} = \alpha(\theta - r) \tag{S9}$$

$$V_{\delta x} = \frac{r(1-r)}{\nu} \tag{S10}$$

α is the return rate to θ. Greater values of α constrain trait variation more tightly around θ. With these coefficients and setting the first constant of integration c_1 to 0 yields:

$$f(r) = c_2 \nu r^{2\alpha\nu\theta - 1}(1-r)^{2\alpha\nu(1-\theta)-1} \tag{S11}$$

where:

$$c_2 = 1 \Big/ \int_0^1 \nu r^{2\alpha\nu\theta-1}(1-r)^{2\alpha\nu(1-\theta)-1} dr \tag{S12}$$

$$= \frac{1}{\nu \mathrm{B}(2\alpha\nu\theta, 2\alpha\nu(1-\theta))} \tag{S13}$$

$\mathrm{B}(\cdot)$ is the beta function. Setting c_1 to 0 can be justified by recognizing that the distribution should be symmetrical ($x = 1 - x$) when $\theta = 0.5$, which only occurs if

$c_1 = 0$ (S.P. Otto pers. comm.). Further, I confirmed the accuracy of the analytically-derived stationary distribution using stochastic simulations (data not shown).

Defining $\phi = \alpha\nu$, the stationary distribution simplifies somewhat to:

$$f(r) = \frac{r^{2\phi\theta-1}(1-r)^{2\phi(1-\theta)-1}}{\mathrm{B}(2\phi\theta, 2\phi(1-\theta))} \tag{S14}$$

This is the Beta(α, β) distribution with $\alpha = 2\phi\theta$ and $\beta = 2\phi(1-\theta)$. Note that, following standard notation, α here refers to the first shape parameter of the Beta distribution, not the constraint factor of the evolutionary process model. This result means that the well-known statistical properties of the Beta distribution can be leveraged to understand the stationary distribution of a proportion trait under a constrained phenotypic landscape. For example, the Beta distribution takes on a variety of shapes that begin to resemble the distribution of proportional traits like stomatal ratio (Fig. S3). Hence, the evolutionary process model developed here provides a strong theoretical justification for fitting the stomatal ratio data to a mixture of Beta distributions in order to infer the regimes shaping this trait across plant species. Although I have derived the model with stomatal ratio in mind, it should be applicable to wide variety of proportional traits evolving under a constrained phenotypic landscapes.

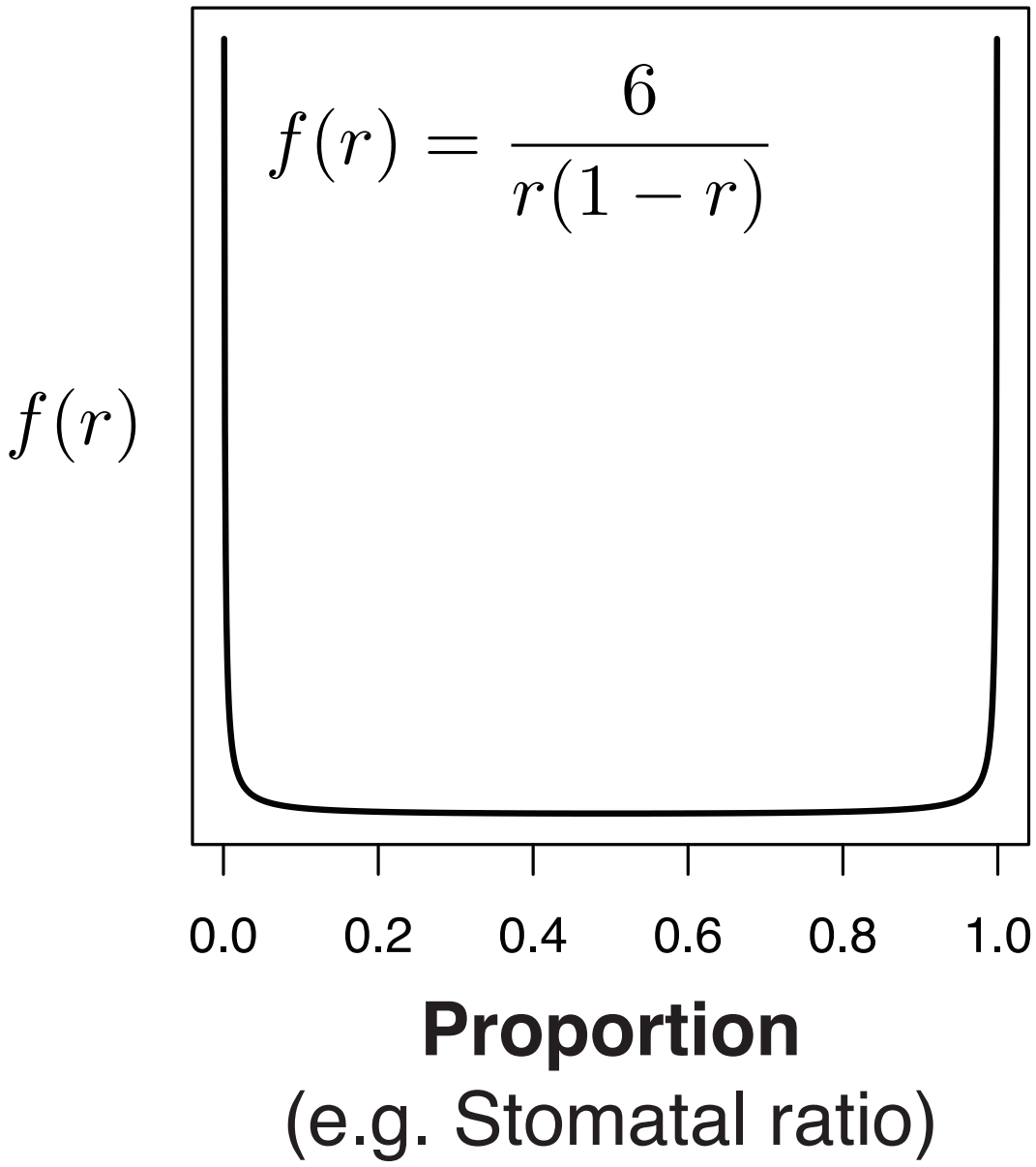

Fig. S2. Without constraint, a proportion trait like stomatal ratio (r) will evolve toward a distribution in which most species are 0 or 1.

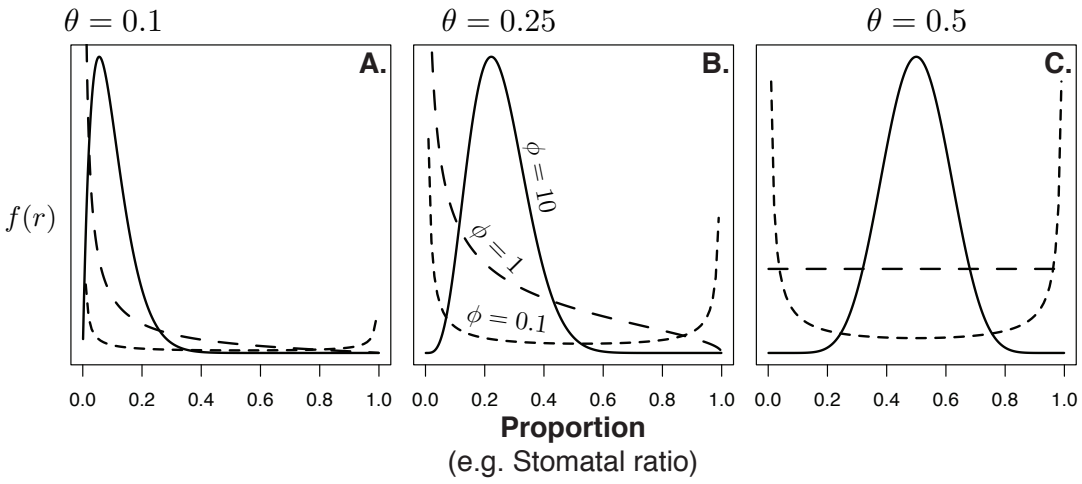

Fig. S3. A proportion trait like stomatal ratio evolving under a single regime is Beta distributed. The Beta distribution can take on a wide variety of shapes depends on the long-run average θ and the levels of constraint ϕ (greater ϕ equals greater constraint).

Appendix S4: Phenotypic distributions have time to reach stationarity

Appendix S3 showed that under a bounded Ornstein-Uhlenbeck model, a proportion trait like stomatal ratio should be Beta-distributed at stationarity. Hence, fitting the data to a Beta distribution would be valid if the trait distribution is close to stationarity. In particular, if the rate (denoted α) at which lineages return to the long-run average (θ) is fast relative to the frequency of regime shifts (e.g. hypo- to amphistomy), then the observed distribution is probably close to stationarity. In this appendix, I estimate the rate of stomatal ratio regime shifts using SIMMAP and the return rate (α) under a standard OU model ($\mathrm{OU}_{\mathrm{norm}}$) as it is not yet possible to estimate this parameter using the $\mathrm{OU}_{\mathrm{beta}}$ model.

Mapping regime shifts using SIMMAP

I calculated the probability that each species (i.e. a tip in the phylogeny) belongs to one of three regimes (hypo, amphi, or intermediate) identified using finite mixture models. The probability of belonging to a given regime was calculated using Eq S21 with parameter estimates from the the three-regime model ($k = 3$ parameters in Table S1). From the tip state probabilities, I estimated the maximum likelihood transition matrix $\boldsymbol{Q}$ between regimes using the make.simmap function in the R package **phytools** version 0.4-56 (58). The function implements SIMMAP, a method for mapping discrete trait evolution on phylogenies (59). The asymmetric transition rate model (AIC = 858) fit better than symmetric (AIC = 878) and equal rate (AIC = 911) models. The full transition matrix is:

$$\boldsymbol{Q} = \begin{pmatrix} & q_{\mathrm{h}\to\mathrm{i}} & q_{\mathrm{h}\to\mathrm{a}} \\ q_{\mathrm{i}\to\mathrm{h}} & & q_{\mathrm{i}\to\mathrm{a}} \\ q_{\mathrm{a}\to\mathrm{h}} & q_{\mathrm{a}\to\mathrm{h}} & \end{pmatrix}$$

The subscripts below each transition rate q indicate the regimes (a = amphistomatous; i = intermediate; h = hypostomatous). The maximum-likelihood estimated transition matrix is:

$$\hat{\boldsymbol{Q}} = \begin{pmatrix} & 0 & 0 \\ 1.8 \times 10^{-2} & & 1.3 \times 10^{-2} \\ 5.0 \times 10^{-4} & 5.1 \times 10^{-3} & \end{pmatrix}$$

Interestingly, the transition rates from the intermediate to hypo- or amphistomatous regimes ($q_{\mathrm{i}\to\mathrm{h}}, q_{\mathrm{i}\to\mathrm{a}}$) are much higher than that from hypostomy or amphistomy to intermediate ($q_{\mathrm{h}\to\mathrm{i}}, q_{\mathrm{a}\to\mathrm{i}}$). This suggests that the intermediate regime is relatively transitory, whereas lineages that enter hypo or amphi regimes remain there for a long time. Next, I simulated 1000 maps of regime shifts on the tree from $\hat{\boldsymbol{Q}}$. For each species, in all 1000 simulated maps, I calculated the median time a species had spent in its current regime. For species in hypo- and amphistomatous regimes, the time was very high, 93 and 225 my, respectively. Species spent much less time in the intermediate regime (median = 42 my).

Estimating phylogenetic signal using OUwie

In Ornstein-Uhlenbeck models, low α indicates a weak 'pull' back toward the long-average θ. In contrast, high α means that traits value bounce around rapidly. Hence, α estimates phylogenetic signal. For example, the phylogenetic half-life, or time it takes for the trait correlation between ancestor and descendant to be halved, is equal to $\log(2)/\alpha$. I simulated 1000 stochastic character maps from the estimated transition matrix $\hat{\boldsymbol{Q}}$. For each tree, I used the R package **OUwie** version 1.45 (60; 61) to fit single and three-regime Brownian motion and Ornstein-Uhlenbeck models. Specifically, I fit single-regime Brownian motion (BM_1) and Ornstein-Uhlenbeck (OU_1) models as well as three-regime Brownian motion (BMS_3 – separate diffusion parameters for each regime) and Ornstein-Uhlenbeck (OUM_3 and $OUMV_3$). Both OUM_3 and $OUMV_3$ models fit different optima (θ) for each regime but assume a shared α. However, the $OUMV_3$ allows different diffusion parameters (σ^2) for each regime. I did not use models ($OUMV_3$, $OUMVA_3$) that estimate different α's for each regime because these were often numerically unstable and the regime-specific α's were very similar (data not shown). Following (62), I estimated parameters as the median across 1000 character maps. I compared model fit using the median BIC across 1000 character maps.

OU_{norm} results confirm those using OU_{beta}: a three-regime OU model fits much better (lower BIC) than single-regime and BM models, even after accounting for phylogenetic nonindependence. It also suggests for hypo and amphi regimes, there is ample time for trait distributions to approach stationarity. The estimated phylogenetic half-life, $log(2)/\alpha = 22$ my is far below the median estimate for time species have evolved in their present-day regime, 42–225 my, depending on the regime.

Table S3: Parameter estimates of one (X_1) and three (X_3) regime models using OUwie. I fit Brownian motion (BM) and Ornstein-Uhlenbeck (OU) models to the stomatal ratio data. Unlike the OU_{beta}parameter estimates, these analyses account for phylogeny. α is the return rate to the long-run average (θ_i) of regime i. σ_i^2 is the diffusion coefficient for regime i. For one regime models, maximum likelihood parameters are reported. For three-regime models, I report the median parameter and Bayesian Information Criterion (BIC) values from 1000 maps of regime shifts across the phylogeny.

Model	α	σ_1^2	σ_2^2	σ_3^2	θ_1	θ_2	θ_3	BIC
BM_1	na	4.9×10^{-4}	na	na	0.162	na	na	-180
OU_1	5.6×10^{-3}	7.3×10^{-4}	na	na	0.188	na	na	-216
BMS_3	na	1.5×10^{-4}	1.5×10^{-3}	2×10^{-4}	0.012	na	na	-462
OUM_3	2.4×10^{-2}	1×10^{-3}	na	na	-0.026	0.238	0.467	-602
$OUMV_3$	3.1×10^{-2}	5×10^{-4}	1.1×10^{-2}	5.6×10^{-4}	-0.005	0.226	0.468	-804

Appendix S5: Fitting evolutionary process to pattern using finite mixture models estimated with maximum likelihood

Before presenting statistical details, I must provide some caveats about my approach. Although I do not think these qualifications affect any of the main conclusions of this paper, they are important for others who might use similar methods or, better yet, seek to extend them. I tested for multiple regimes using a conceptually similar but somewhat different approach than previous studies. Current methods for inferring multiple selective regimes (54; 55; 56) cannot yet accommodate Beta-distributed traits, and there is no general solution to the stochastic differential equation in Appendix S3. Future work is needed to develop numerical methods to integrate the bounded Ornstein-Uhlenbeck process model elaborated here into existing statistical frameworks for multi-regime inference. However, the fact that distantly-related plant families converge on similar trait distributions strong suggests that the main results are robust to phylogenetic nonindependence.

I infer the number of regimes acting on stomatal ratio by fitting a mixture of stationary distributions derived from the process model above to the data. By fitting the data to the stationary distribution, I implicitly assume that evolution is sufficiently rapid to ignore phylogenetic signal. Numerical simulations of the diffusion indicate that the transitory distribution is also Beta (data not shown), meaning that evidence for multiple regimes cannot be an artifact of transitory behaviour within a single regime. In this section I derive the likelihood functions and describe an expectation-maximization algorithm to find the maximum likelihood mixture model

given the data. R code to implement these methods is available on Dryad (50). In general, finite mixture distributions are the summation of $k \geq 2$ mixture components (i.e. probability distributions) with density $f_i(x)$ and mixture weight w_i:

$$g(x; k) = \sum_{i=1}^{k} w_i f_i(x) \tag{S15}$$

Here the i-th mixture component has a probability density $f_i(x)$ given by the stationary distribution in Eq S14 with parameters θ_i, ϕ_i. The likelihood of a mixture distribution given k mixture components and a data vector $\boldsymbol{x}$ with sample size n is the weighted sum of the likelihoods of each component:

$$\mathcal{L}(\boldsymbol{w}, \boldsymbol{\phi}, \boldsymbol{\theta}; \boldsymbol{x}, k) = \sum_{i=1}^{k} w_i \mathcal{L}_i(\phi_i, \theta_i; \boldsymbol{x}) \tag{S16}$$

The parameter vectors $\boldsymbol{w}$, $\boldsymbol{\phi}$, and $\boldsymbol{\theta}$ are defined as:

$$\boldsymbol{w} := \{w_1, \ldots, w_k\} \tag{S17}$$

$$\boldsymbol{\phi} := \{\phi_1, \ldots, \phi_k\} \tag{S18}$$

$$\boldsymbol{\theta} := \{\theta_1, \ldots, \theta_k\} \tag{S19}$$

For the i-th component, the likelihood of parameters ϕ_i and θ_i given the data is the

product of the probability densities of each datum $(x_1, x_2, \ldots, x_n)$:

$$\mathcal{L}_i(\phi_i, \theta_i; \boldsymbol{x}) = \prod_{j=1}^{n} f_i(x_j; \phi_i, \theta_i) \tag{S20}$$

To obtain reasonable fits, I found it necessary to modify the likelihood to incorporate left- and right-censored data. This is because the stomatal ratio dataset contains many 0's (all stomata are on the lower surface of the leaf) and 1's (all stomata on the upper surface). Under most parameterizations of the Beta distribution, the probability density of 0 and 1 is ∞ or 0. I left- and right-censored the data at $x_l = 0.001$ and $x_r = 0.999$ as these were very close to the lowest and highest values reported in the dataset (except 0 and 1), respectively. This means that any datum reported as 0 was statistically interpreted as falling anywhere between 0 and 0.001. Likewise, a datum reported as 1 was assumed to fall between 0.999 and 1. A reasonable interpretation is that a stomatal ratio so close to 0 or 1 would be practically difficult to measure. Biologically, a stomatal ratio less than 0.001 or greater than 0.999 are indistinguishable from 0 and 1. With censoring, the likelihood of the i-th component becomes:

$$\mathcal{L}_i(\phi_i, \theta_i; \boldsymbol{x}) = \prod_{j=1}^{n} f(x; \phi_i, \theta_i)^{I_l(x) I_r(x)} F(x_l; \phi_i, \theta_i)^{1 - I_l(x)} (1 - F(x_r; \phi_i, \theta_i))^{1 - I_r(x)} \tag{S21}$$

$F(x; \phi_i, \theta_i)$ is the cumulative density function of the Beta distribution; $I_l(x)$ and $I_r(x)$ are indicator functions:

$$I_l(x) = \begin{cases} 0 & \text{if } x = x_l \\ 1 & \text{if } x \neq x_l \end{cases} \tag{S22}$$

$$I_r(x) = \begin{cases} 0 & \text{if } x = x_r \\ 1 & \text{if } x \neq x_r \end{cases} \tag{S23}$$

To find the maximum likelihood mixture distribution, I used an expectation-maximization (EM) algorithm similar to (63). EM algorithms are particularly well-suited to fitting mixture distributions. Here, I describe the initialization, expectation (E-step), and maximization (M-step) procedure.

Initialization

The data were divided into k evenly-sized components. For example, if $k = 2$, data below the median were assigned to component 1; data above the median were assigned to component 2. For each component, the initial weight was therefore $w_{i,\text{init}} = 1/k$. Within each component, I used the optim function in R to estimate the maximum likelihood parameters ($\hat{\phi}_i^{(\text{init})}$ and $\hat{\theta}_i^{(\text{init})}$) of a Beta distribution. Note that I am using parenthetical superscript to indicate the iteration of the algorithm, starting with the initial parameterization, followed by $t = 1, 2, 3, \ldots$ until the likelihood converges. The initial parameter vectors are therefore:

$$\boldsymbol{w}^{(\text{init})} := \{1/k, \ldots, 1/k\} \tag{S24}$$

$$\boldsymbol{\phi}^{(\text{init})} := \{\hat{\phi}_1^{(\text{init})}, \ldots, \hat{\phi}_k^{(\text{init})}\} \tag{S25}$$

$$\boldsymbol{\theta}^{(\text{init})} := \{\hat{\theta}_1^{(\text{init})}, \ldots, \hat{\theta}_k^{(\text{init})}\} \tag{S26}$$

Expectation

In the E-step, the expected likelihood is calculated under the parameters estimated from the previous iteration. The mixture weights are then updated and carried forward to the M-step. For the first iteration following initialization, the mixture weights $\boldsymbol{w}^{(1)}$ conditional on the initial parameterization are:

$$w_i^{(1)} = \frac{\sum_{j=1}^{N} y_{ij}^{(\text{init})}}{n} \tag{S27}$$

where $y_{ij}^{(\text{init})}$ is the probability that x_j belongs to component i given initial parameters:

$$y_{ij}^{(\text{init})} = \frac{w_i^{(\text{init})} f(x_j; \hat{\phi}_i^{(\text{init})}, \hat{\theta}_i^{(\text{init})})}{g(x_j; k, \boldsymbol{w}^{(\text{init})}, \boldsymbol{\phi}^{(\text{init})}, \boldsymbol{\theta}^{(\text{init})})} \tag{S28}$$

In subsequent iterations, the equations are similarly:

$$w_i^{(t+1)} = \frac{\sum_{j=1}^{N} y_{ij}^{(t)}}{N} \tag{S29}$$

$$y_{ij}^{(t)} = \frac{w_i^{(t)} f(x_j, \phi_i^{(t)}, \theta_i^{(t)})}{g(x_j; k, \boldsymbol{w}^{(t)}, \boldsymbol{\phi}^{(t)}, \boldsymbol{\theta}^{(t)})} \tag{S30}$$

Maximization

During the M-step, estimates of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are updated using maximum likelihood conditional on mixture weights calculated in the E-step:

$$\{\boldsymbol{\phi}^{(t+1)}, \boldsymbol{\theta}^{(t+1)}\} = \arg\max_{\boldsymbol{\phi}, \boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\theta}; \boldsymbol{x}, k, \boldsymbol{w}^{(t)}) \tag{S31}$$

I used the optim function in R to find the parameters that maximized the likelihood function. After the M-step, the next iteration begins at the E-step and continues until the likelihood converges to a stable value. As with other hill-climbing likelihood searches, EM does not guarantee convergence at the maximum likelihood. With the stomatal ratio data, I found that multiple initialization procedures yielded the same final parameter estimates, suggesting that the algorithm was successfully converging on the maximum likelihood solution.

Appendix S6: Cost-benefit model

In this model, I opted to tradeoff the precision of a biophysical diffusion model for a more general, albeit realistic, model with fewer parameters. Hence, the cost-benefit model of stomatal ratio is true to the underlying physics but otherwise not strongly dependent on specific assumptions. Future work will be needed to test if this more general model is consistent with mechanistic biophysical models. The symbols used in the model are summarized in Table S5.

Table S4. Glossary of symbols used in the cost-benefit model.

Symbol	Description
SR	Stomatal ratio: ratio of upper to total stomatal density
S	logit of stomatal ratio (SR)
S_{opt}	Stomatal ratio (logit scale) that maximizes fitness benefits
B_{max}	Maximum fitness benefit when $S = S_{\text{opt}}$
σ^2	Shape factor of benefit function
C_{max}	Maximum fitness cost of when all stomata are on the upper side ($SR = 1$)
S_{fit}	Stomatal ratio maximizes fitness (benefits minus costs)

I model selection on the logit of stomatal ratio (upper:total), which I denote $S = \text{logit}(SR) = \log\big(SR/(1-SR)\big)$, so that feasible trait variation (SR is constrained from 0 to 1) is continuous and unbounded. Fitness as a function of stomatal ratio depends on the difference between the benefits ($f(S)$) minus the costs ($g(S)$). Therefore, fitness as a function of stomatal ratio is:

$$W(S) = 1 + f(S) - g(S) \tag{S32}$$

Based on biophysical theory (1; 2), I assume that there is an intermediate optimal

stomatal ratio (S_{opt}) at which photosynthetic rate is maximized. Above and below that optimum, photosynthetic rate decreases, which I modelled as a Gaussian function:

$$f(S) = B_{\text{max}} e^{-\frac{(S-S_{\text{opt}})^2}{2\sigma^2}} \tag{S33}$$

B_{max} defines the maximum fitness when $S = S_{\text{opt}}$. B_{max} is a complex function of the external environment, total stomatal conductance, internal photosynthetic capacity, and other factors. For simplicity, I do not explicitly model how these factors affect B_{max} here, but rather treat it is a phenomenological variable. σ^2 acts akin to a shape factor when the function is viewed from a logit scale. When σ^2 is large, the benefit function has an inverted-U shape. There are increasing returns to fitness of the first few upper stomata, but diminishing returns to further increases in S (Fig. 3A). In contrast, when σ^2 is small, the benefit function is more bell-shaped; the fitness benefit of the first few upper stomata is large, but with diminishing returns (Fig. 3C).

I assumed a linear cost (e.g. increased susceptibility to foliar pathogens (32)) for each additional upper stomate. Note however that the model is agnostic to the specific mechanism underlying the fitness cost or costs. The total cost as a function of stomatal ratio is the product of the total stomatal density, the stomatal ratio (upper:total density), and the cost per upper stomate. I define the slope of the cost function as C_{max}, which is equal to the total stomatal density times the cost per upper stomate:

$$h(SR) = C_{\max} SR \tag{S34}$$

401 On a logit scale, the total cost asymptotically approaches $C_{\max}$:

$$g(S) = \frac{C_{\max}}{1 + e^{-S}} \tag{S35}$$

402 If more were known about the cost of having upper stomata, a more realistic model
403 could be constructed. Without such knowledge, I believe it is judicious to start with
404 the simplest model that makes few assumptions and therefore could apply to a large
405 number of particular underlying mechanisms. Substituting Eqs S33 and S35 into
406 Eq S32, fitness as a function of S is:

$$W(S) = 1 + B_{\max} e^{-\frac{(S - S_{\text{opt}})^2}{2\sigma^2}} - \frac{C_{\max}}{1 + e^{-S}} \tag{S36}$$

407 Note that if the cost function were applied to lower rather than upper stomata, as
408 might be the case for specialized taxa such as aquatic plants, then one could obtain
409 the same results, except that hyper- rather than hypostomy would prevail, as in the
410 Poaceae data. The fitness function is maximized where the marginal benefit of the
411 next upper stomate is equal to the marginal cost:

$$\frac{df(S)}{dS} = \frac{dg(S)}{dS} \tag{S37}$$

412 I did not obtain an analytical solution, so instead I used the optim function in R

(64) to numerically solve for the stomatal ratio that maximized fitness (S_{fit}) under varying ratios of fitness cost (C_{max}) to benefit (B_{max}). I tuned the benefit:cost ratio by fixing B_{max} to 1 and varying C_{max} between 0.01 and 100. I also varied the shape factor σ^2 between 0.1 and 10, which appeared to capture the full range of relevant model behaviour. For all numerical solutions, I assumed that the optimal stomatal ratio for photosynthesis was 0.5, hence $S_{\text{opt}} = 0$ on a logit scale. Next, I generated hypothetical trait distributions under a scenario where the benefit:cost ratio varies uniformly from 10^{-2} to 10^{2}. I solved for S_{fit} with 10^4 evenly spaced values of $B_{\text{max}} : C_{\text{max}}$ under low, medium, and high values of σ^2. R code for finding numerical solutions is available from Dryad (50).

Appendix S7: Data Sources

1. Boeger and Gluzezak 2006
2. Brodribb *et al.* 2013
3. Camargo and Marenco 2011
4. Cooper and Cass 2003 ; Cooper *et al.* 2004
5. Dickie and Gasson 1999
6. Dunbar-Co *et al.* 2009
7. Fahmy 1997
8. Fahmy *et al.* 2007
9. Fontenelle *et al.* 1994
10. Giuliani *et al.* 2013
11. Holbrook and Putz 1996
12. Körner *et al.* 1989
13. Lohr 1919
14. Loranger and Shipley 2010
15. Malaisse and Colonval-Elenkov 1982
16. Maricle *et al.* 2009

17. Muir *et al.* 2014

18. Parkin and Pearson 1903

19. Peace and MacDonald 1981

20. Rao and Tan 1980

21. Reed *et al.* 2000

22. Ridge *et al.* 1984

23. Selvi and Bigazzi 2001

24. Seshavatharam and Srivalli 1989

25. Sobrado and Medina 1980

Appendix S8: Additional detail on stomatal ratio distributions within families

There were at least two selective regimes in 9 of 10 families analyzed (Fig. 2). In one family, Rubiaceae, all species were inferred as members of a hypostomatous regime. Two regimes are supported in most (8 of 9) multi-regime families, except Asteraceae, in which three regimes are favoured (Fig. 2A).In all mutli-regime families except Poaceae, there are distinct regimes associated with hypo- and amphistomy; in Poaceae, there are hyper- and amphistomous regimes instead (Fig. 2E). However, the hyperstomatous species of Poaceae in this study may not be representative of the family since they are wetland specialists in the genus *Spartina* (77). Generally, the internal (i.e. amphistomatous) mode is closely centered around 0.5, as predicted from biophysical theory (1; 2), except in in the Rosaceae, where the inferred optimum is closer to 0.25.

SI References

[1] Parkhurst, D. F., 1978 The adaptive significance of stomatal occurrence on one or both surfaces of leaves. *The Journal of Ecology* **66**, 367–383.

[2] Gutschick, V. P., 1984 Photosynthesis model for C_3 leaves incorporating CO_2 transport, propagation of radiation, and biochemistry 2. ecological and agricultural utility. *Photosynthetica* **18**, 569–595.

[3] Parkhurst, D. F. & Mott, K. A., 1990 Intercellular diffusion limits to CO_2 uptake in leaves studied in air and helox. *Plant Physiology* **94**, 1024–1032.

[4] Smith, W. K., Bell, D. T. & Shepherd, K. A., 1998 Associations between leaf structure, orientation, and sunlight exposure in five Western Australian communities. *American Journal of Botany* **85**, 51–63.

[5] James, S. A. & Bell, D. T., 2001 Leaf morphological and anatomical characteristics of heteroblastic *Eucalyptus globulus* ssp. *globulus* (myrtaceae). *Australian Journal of Botany* **49**, 259–269.

[6] Burrows, G., 2001 Comparative anatomy of the photosynthetic organs of 39 xeromorphic species from subhumid New South Wales, Australia. *International Journal of Plant Sciences* **162**, 411–430.

[7] Williams, M., Woodward, F. I., Baldocchi, D. D. & Ellsworth, D. S., 2004 CO_2 capture: Leaf to landscape. In *Photosynthetic Adaptation* (eds. W. K. Smith, T. C. Vogelmann & C. Critchley), pp. 133–168. New York: Springer.

[8] Brodribb, T. J., Jordan, G. J. & Carpenter, R. J., 2013 Unified changes in cell

size permit coordinated leaf evolution. *New Phytologist* **199**, 559–570.

[9] Muir, C. D., Hangarter, R. P., Moyle, L. C. & Davis, P. A., 2014 Morphological and anatomical determinants of mesophyll conductance in wild relatives of tomato (*Solanum* sect. *Lycopersicon*, sect. *Lycopersicoides*; Solanaceae). *Plant, Cell & Environment* **37**, 1415–1426.

[10] Goble-Garratt, E., Bell, D. & Loneragan, W., 1981 Floristic and leaf structure patterns along a shallow elevational gradient. *Australian Journal of Botany* **29**, 329–347.

[11] Mott, K. A., Gibson, A. C. & O'Leary, J. W., 1984 The adaptive significance of amphistomatic leaves. *Plant, Cell & Environment* **5**, 455–460.

[12] Beerling, D. J. & Kelly, C. K., 1996 Evolutionary comparative analyses of the relationship between leaf structure and function. *New Phytologist* **134**, 35–51.

[13] Poorter, H., Niinemets, Ü., Walter, A., Fiorani, F. & Schurr, U., 2010 A method to construct dose–response curves for a wide range of environmental factors and plant traits by means of a meta-analysis of phenotypic data. *Journal of Experimental Botany* **61**, 2043–2055.

[14] Gibson, A. C., 1996 *Structure-Function Relations of Warm Desert Plants*. Berlin: Springer-Verlag.

[15] Lohr, P. L., 1919 *Untersuchungen über die Blattantomie von Alphen- und Ebenenpflanzen*. Ph.D. thesis, Universität Basel.

[16] Fitter, A. & Peat, H., 1994 The ecological flora database. *Journal of Ecology* **82**, 415–425.

[17] Jordan, G. J., Carpenter, R. J. & Brodribb, T. J., 2014 Using fossil leaves as evidence for open vegetation. *Palaeogeography, Palaeoclimatology, Palaeoecology* **395**, 168–175.

[18] Wood, J. G., 1934 The physiology of xerophytism in Australian plants: the stomatal frequencies, transpiration and osmotic pressures of sclerophyll and tomentose-succulent leaved plants. *Journal of Ecology* **22**, 69–87.

[19] Mott, K. A. & O'Leary, J. W., 1984 Stomatal behavior and CO_2 exchange characteristics in amphistomatous leaves. *Plant physiology* **74**, 47–51.

[20] Spinner, H., 1936 Stomates et altitude. *Berichte der Schweizerischen Botanischen Gesellschaft* **46**, 12–27.

[21] Woodward, F. I., 1986 Ecophysiological studies on the shrub *Vaccinium myrtillus* l. taken from a wide altitudinal range. *Oecologia* **70**, 580–586.

[22] Körner, C., Neumayer, M., Menendez-Riedl, S. P. & Smeets-Scheel, A., 1989 Functional morphology of mountain plants. *Flora* **182**, 353–383.

[23] Vogel, S., 2012 *The Life of a Leaf*. Chicago: University of Chicago Press.

[24] Metcalfe, C. R. & Chalk, L., 1950 *Anatomy of the dicotyledons, Vols. 1 & 2*. Oxford: Oxford University Press, first edition.

[25] Franks, P. J. & Beerling, D. J., 2009 Maximum leaf conductance driven by CO_2 effects on stomatal size and density over geologic time. *Proceedings of the National Academy of Sciences* **106**, 10343–10347.

525 [26] Foster, J. & Smith, W., 1986 Influence of stomatal distribution on transpiration
526 in low-wind environments. *Plant, Cell & Environment* **9**, 751–759.

527 [27] Gates, F. C., 1914 Winter as a factor in the xerophily of certain evergreen
528 ericads. *Botanical Gazette* **57**, 445–489.

529 [28] Pospíšilová, J. & Solárová, J., 1984 Environmental and biological control of
530 diffusive conductances of adaxial and abaxial leaf epidermes. *Photosynthetica*
531 **18**, 445–453.

532 [29] Smith, W., 1981 Temperature and water relation patterns in subalpine under-
533 story plants. *Oecologia* **48**, 353–359.

534 [30] Reich, P., 1984 Relationships between leaf age, irradiance, leaf conductance,
535 CO_2 exchange, and water-use efficiency in hybrid poplar. *Photosynthetica* **18**,
536 445–453.

537 [31] Smith, W. K. & McClean, T. M., 1989 Adaptive relationship between leaf wa-
538 ter repellency, stomatal distribution, and gas exchange. *American Journal of*
539 *Botany* **76**, 465–469.

540 [32] McKown, A. D., Guy, R. D., Quamme, L., Klápště, J., La Mantia, J., Constabel,
541 C., El-Kassaby, Y. A., Hamelin, R. C., Zifkin, M. & Azam, M., 2014 Association
542 genetics, geography and ecophysiology link stomatal patterning in *Populus tri-*
543 *chocarpa* with carbon gain and disease resistance trade-offs. *Molecular Ecology*
544 **23**, 5771–5790.

545 [33] Napp-Zinn, K., 1970 *Anatomie des blattes*. Berlin: Borntraeger.

[34] Metcalfe, C. R. & Chalk, L., 1979 *Anatomy of the dicotyledons, Vols. 1 & 2.* Oxford: Clarendon Press, second edition.

[35] Abràmoff, M. D., Magalhães, P. J. & Ram, S. J., 2004 Image processing with imagej. *Biophotonics International* **11**, 36–42.

[36] Leick, E., 1927 Untersuchungen über den Einfluß des Lichtes auf die öffnungsweite unterseitiger und oberseitiger Stomata desselben Blattes. *Jahrbücher für Wissenschaftliche Botanik* **67**, 771–848.

[37] Salisbury, E., 1927 On the causes and ecological significance of stomatal frequency, with special reference to the woodland flora. *Philosophical Transactions of the Royal Society of London. Series B* **216**, 1–65.

[38] Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A., 2005 Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965–1978.

[39] Los, S. O., 2010 ISLSCP II FASIR-adjusted NDVI, 1982-1998. volume 10.

[40] Chamberlain, S., Ram, K., Barve, V. & Mcglinn, D., 2014 *rgbif: Interface to the Global Biodiversity Information Facility API*. R package version 0.7.8.99.

[41] USDA, NRCS, 2014. The plants database. Available: http://www.plants.usda.gov.

[42] Encyclopedia of Life, 2015. Available at: http://www.eol.org.

[43] Boyle, B., Hopkins, N., Lu, Z., Garay, J. A. R., Mozzherin, D., Rees, T., Matasci, N., Narro, M. L., Piel, W. H., Mckay, S. J. *et al.*, 2013 The taxonomic name

resolution service: an online tool for automated standardization of plant names. *BMC bioinformatics* **14**, 16.

[44] The Plant List, 2013. Version 1.1. Available: http://www.theplantlist.org/.

[45] The International Plant Names Index, 2015. Available at: http://www.ipni.org.

[46] Euro+Med Plantbase, 2015. Available at: http://www.emplantbase.org.

[47] Giuliani, R., Koteyeva, N., Voznesenskaya, E., Evans, M. A., Cousins, A. B. & Edwards, G. E., 2013 Coordination of leaf photosynthesis, transpiration, and structural traits in rice and wild relatives (genus *Oryza*). *Plant Physiology* **162**, 1632–1651.

[48] Webb, C. O. & Donoghue, M. J., 2005 Phylomatic: tree assembly for applied phylogenetics. *Molecular Ecology Notes* **5**, 181–183.

[49] Orme, C. D. L., Freckleton, R. P., Thomas, G. H., Petzoldt, T., Fritz, S. A., Isaac, N. J. B. & Pearse, W. D., 2013 *caper: Comparative Analyses of Phylogenetics and Evolution in R.* R package version 0.5.2.

[50] Muir, C. D., 2015 *Data from: Making pore choices: repeated regime shifts in stomatal ratio. Dryad Digital Repository.*

[51] Hansen, T. F., 1997 Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**, 1341–1351.

[52] Hansen, T. F., 2012 Adaptive landscapes and the comparative analysis of adaptation. In *The Adaptive Landscape in Evolutionary Biology* (eds. E. Svensson & R. Calsbeek), pp. 205–226. Oxford, UK: Oxford University Press.

588 [53] Pennell, M. W. & Harmon, L. J., 2013 An integrative view of phylogenetic
589 comparative methods: connections to population genetics, community ecology,
590 and paleobiology. *Annals of the New York Academy of Sciences* **1289**, 90–105.

591 [54] Butler, M. & King, A. A., 2004 Phylogenetic comparative analysis: a modeling
592 approach for adaptive evolution. *The American Naturalist* **164**, 683–695.

593 [55] Ingram, T. & Mahler, D. L., 2013 SURFACE: detecting convergent evolu-
594 tion from comparative data by fitting Ornstein-Uhlenbeck models with stepwise
595 Akaike Information Criterion. *Methods in Ecology and Evolution* **4**, 416–425.

596 [56] Uyeda, J. C. & Harmon, L. J., 2014 A novel Bayesian method for inferring and
597 interpreting the dynamics of adaptive landscapes from phylogenetic comparative
598 data. *Systematic Biology* **64**, 902–918.

599 [57] Otto, S. P. & Day, T., 2007 *A Biologist's Guide to Mathematical Modeling in*
600 *Ecology and Evolution*. Princeton, New Jersey: Princeton University Press.

601 [58] Revell, L. J., 2012 phytools: An R package for phylogenetic comparative biology
602 (and other things). *Methods in Ecology and Evolution* **3**, 217–223.

603 [59] Huelsenbeck, J. P., Nielsen, R. & Bollback, J. P., 2003 Stochastic mapping of
604 morphological characters. *Systematic Biology* **52**, 131–158.

605 [60] Beaulieu, J. M., Jhwueng, D.-C., Boettiger, C. & OâĂŹMeara, B. C., 2012 Mod-
606 eling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive
607 evolution. *Evolution* **66**, 2369–2383.

608 [61] Beaulieu, J. M. & O'Meara, B., 2015 *OUwie: Analysis of Evolutionary Rates in*
609 *an OU Framework*. R package version 1.45.

610 [62] Bossu, C. M. & Near, T. J., 2015 Ecological constraint and the evolution of
611 sexual dichromatism in darters. *Evolution* **69**, 1219–1231.

612 [63] Grün, B., Kosmidis, I. & Zeileis, A., 2012 Extended beta regression in R: Shaken,
613 stirred, mixed, and partitioned. *Journal of Statistical Software* **48**, 1–25.

614 [64] R Core Team, 2013 *R: A Language and Environment for Statistical Computing*.
615 R Foundation for Statistical Computing, Vienna, Austria.

616 [65] Boeger, M. R. T. & Gluzezak, R. M., 2006 Adaptações estruturais de sete espé-
617 cies de plantas para as condições ambientais da área de dunas de Santa Catarina,
618 Brasil. *Iheringia, Série Botânica* **61**, 73–82.

619 [66] Camargo, M. A. B. & Marenco, R. A., 2011 Density, size and distribution of
620 stomata in 35 rainforest tree species in Central Amazonia. *Acta Amazonica* **41**,
621 205–212.

622 [67] Cooper, R. L. & Cass, D. D., 2003 A comparative epidermis study of the
623 Athabasca sand dune willows (*Salix*; Salicaceae) and their putative progenitors.
624 *Canadian Journal of Botany* **81**, 749–754.

625 [68] Cooper, R. L., Ware, J. V. & Cass, D. D., 2004 Leaf thickness of *Salix* spp.
626 (Salicaceae) from the Athabasca sand dunes of northern Saskatchewan, Canada.
627 *Canadian Journal of Botany* **82**, 1682–1686.

628 [69] Dickie, J. & Gasson, P., 1999 Comparative leaf anatomy of the Penaeaceae
629 and its ecological implications. *Botanical Journal of the Linnean Society* **131**,
630 327–351.

631 [70] Dunbar-Co, S., Sporck, M. J. & Sack, L., 2009 Leaf trait diversification and

design in seven rare taxa of the Hawaiian *Plantago* radiation. *International Journal of Plant Sciences* **170**, 61–75.

[71] Fahmy, G. M., 1997 Leaf anatomy and its relation to the ecophysiology of some non-succulent desert plants from Egypt. *Journal of Arid Environments* **36**, 499–526.

[72] Fahmy, G. M., Hegazy, A. K., Ali, M. I. A. & Gomaa, N. H., 2007 Structure-function relations of leaves in 24 species of winter and summer weeds. *Global Journal of Environmental Research* **1**, 103–116.

[73] Fontenelle, G., Costa, C. & Machado, R., 1994 Foliar anatomy and micromorphology of eleven species of *Eugenia* L.(Myrtaceae). *Botanical Journal of the Linnean Society* **116**, 111–133.

[74] Holbrook, N. M. & Putz, F., 1996 From epiphyte to tree: differences in leaf structure and leaf water relations associated with the transition in growth form in eight species of hemiepiphytes. *Plant, Cell & Environment* **19**, 631–642.

[75] Loranger, J. & Shipley, B., 2010 Interspecific covariation between stomatal density and other functional leaf traits in a local flora. *Botany* **88**, 30–38.

[76] Malaisse, F. & Colonval-Elenkov, E., 1982 On the leaf anatomy of trees and shrubs of the montane evergreen foresst of Malawi and Zimbabwe. *Geo-Eco-Trop* **6**, 139–160.

[77] Maricle, B. R., Koteyeva, N. K., Voznesenskaya, E. V., Thomasson, J. R. & Edwards, G. E., 2009 Diversity in leaf anatomy, and stomatal distribution and

conductance, between salt marsh and freshwater species in the C_4 genus *Spartina* (Poaceae). *New Phytologist* **184**, 216–233.

[78] Parkin, J. & Pearson, H. H. W., 1903 The botany of the Ceylon patanas. *Journal of the Linnean Society of London, Botany* **35**, 430–463.

[79] Peace, W. & Macdonald, F., 1981 An investigation of the leaf anatomy, foliar mineral levels, and water relations of trees of a Sarawak forest. *Biotropica* **13**, 100–109.

[80] Rao, A. N. & Tan, H., 1980 Leaf structure and its ecological significance in certain mangrove plants. In *Proceedings of the Asian Symposium on Mangrove Environment: Research and Management* (eds. E. Soepadmo, A. N. Rap & D. J. MacIntosh), pp. 183–194. Ardyas.

[81] Read, J., Edwards, C., Sanson, G. D. & Aranwela, N., 2000 Relationships between sclerophylly, leaf biomechanical properties and leaf anatomy in some Australian heath and forest species. *Plant Biosystems* **134**, 261–277.

[82] Ridge, R., Loneragan, W., Bell, D., Colquhoun, I. & Kuo, J., 1984 Comparative studies in selected species of *Eucalyptus* used in rehabilitation of the northern Jarrah forest, Western Australia. ii. wood and leaf anatomy. *Australian Journal of Botany* **32**, 375–386.

[83] Selvi, F. & Bigazzi, M., 2001 Leaf surface and anatomy in Boraginaceae tribe Boragineae with respect to ecology and taxonomy. *Flora* **196**, 269–285.

[84] Seshavatharam, V. & Srivalli, M., 1989 Systematic leaf anatomy of some Indian

674 mangroves. *Proceedings of the Indian Academy of Sciences: Plant Sciences* **99**,
675 557–565.

676 [85] Sobrado, M. & Medina, E., 1980 General morphology, anatomical structure,
677 and nutrient content of sclerophyllous leaves of the âĂŸBanaâĂŹ vegetation of
678 Amazonas. *Oecologia* **45**, 341–345.