**Global biogeography of mating system variation in seed plants**

**APPENDICES**

Table of Contents

Appendix S1: Distribution of mating system variation 2-6

Appendix S2: OLS and phylogenetic regression models for outcrossing rate 7

Appendix S3: Results of F tests from OLS regressions 8

Appendix S4: Estimates of coefficients for outcrossing rate 9-10

Appendix S5: Biogeographic analyses within life history and growth form 11

Appendix S6: Logistic regression and phylogenetic logistic regression models for SI 12

Appendix S7: Logistic regression analyses for the probability of SI 13

Appendix S8: Estimates of coefficients for SI 14-15

Appendix S9: Mulitnomial logistic regressions of life history and growth form 16-19

Appendix S10: Tests of phylogenetic signal in latitude and outcrossing rate 20

Appendix S11: Phylogenetic logistic regressions for the probability of SC 21

**Appendix S1.** Distribution of mating system variation

**Introduction**

The distribution of outcrossing rates in seed plants has been of considerable interest because of theoretical studies predicting that only alternative extreme mating systems are stable (Lande & Schemske 1985) versus that intermediate outcrossing rates are sometimes stable (Johnston 1998; Vallejo-Marin & Uyenoyama 2004; Porcher & Lande 2005; Johnston et al. 2009). A number of previous papers have evaluated the shape of the distribution of mating system variation, along with exploring potential biases in the dataset (Schemske & Lande 1985; Vogler & Kalisz 2001; Goodwillie et al. 2005; Igic & Kohn 2006; Goodwillie et al. 2010; Raduski et al. 2012). Here, we revisited the distribution of mating system variation using a substantially larger data set than previously analyzed. Our new analyses examine if and how the distribution has changed over the 57 years of study. We also test whether the distribution of outcrossing rates is bimodal or includes additional modes.

**Methods**

We tested whether the distribution of outcrossing rates differed significantly from normality using a Shapiro-Wilk test. We also used Gaussian mixture models to test whether the distribution was best fit by single or multiple normal distributions; mixture models examine whether the observed distribution is drawn from a single population or more than one subpopulation. We tested for differences between models that allowed for one through five subpopulations using likelihood ratio tests. Given that the outcrossing rate dataset has expanded substantially over the 57 years of study, we also tested whether the distribution of outcrossing rates has shifted over time and the potential causes of a change in the distribution. We first examined the relationship between outcrossing rate estimates and year. Second, we tested for differences between the current distribution and subsets of the dataset including studies published through 1980, 1985 (Schemske & Lande), 1990, and 2000 using nonparametric Kolmogorov-Smirnov tests.

**Results & Discussion**

The distribution of outcrossing rates differed significantly from a single normal distribution (W = 0.89, p < 0.0001). The distribution was best fit by a four-distribution mixture model, with means of the distributions at 0.04, 0.35, 0.67, and 0.92 (Table S1.1). We found a significant increase in the mean tm among published studies from 1956 to 2012 (F1, 617 = 15.1, P < 0.0001). Cumulative distribution plots show a progressive shift in distribution from 1980 to 2012 as data have accumulated from 47 to 492 taxa (Fig. S1.2). The distribution from 2012 (n = 492) differed significantly from that reported originally in 1985 (n = 55; two-sample Kolmogorov-Smirnoff test: P < 0.0001). Regressions of the percentage of taxa categorized as highly selfing (tm 0.2), outcrossing (tm > 0.8), or mixed mating (0.2 < tm 0.8) on date show that the fraction of selfing species has significantly declined (F1, 5  = 28.4, P < 0.01), the fraction of outcrossers has significantly increased (F1, 5 = 22.8, P < 0.01), and the fraction of species with mixed mating systems has significantly increased (F1, 5 = 9.9, P < 0.05). As expected, tm was significantly greater for SI compared to SC species in our dataset (means: SI = 0.86, SC = 0.60; t = 7.4, P < 0.0001).

Our dataset adds to the growing body of empirical (Barrett & Harder 1996; Vogler & Kalisz 2001; Barrett 2003; Goodwillie et al.2005; Eckert et al. 2006; Goodwille et al. 2010) and theoretical evidence (e.g. Johnston 1998; Vallejo-Marin & Uyenoyama 2004; Porcher & Lande 2005; Johnston et al. 2009) that mixed mating is common. Our analyses show that the fraction of highly outcrossing and mixed mating species (especially those that are primarily outcrossing) in the dataset has increased substantially through time; both of these mating system strategies have been shown to be under-sampled in previous analyses (Igic & Kohn 2006; Goodwillie et al. 2010; Raduscki et al. 2012). In addition, we identified four modes (independent Gaussian distributions) in the distribution of outcrossing rates. Two of these modes are at the extremes of the distribution where species are highly outcrossing (mostly SI species; mode of tm = 0.92) and highly selfing (mode = 0.04). A third mode includes primarily selfing species that reproduce occasionally via outcrossing (mode = 0.35) and a fourth mode includes a large number of (self-compatible) species that primarily outcross but with considerable frequency of selfing (mode = 0.67). In the more highly outcrossing mixed mating species (mode = 0.67), large floral displays may have evolved to attract animal pollinators but (geitonogamous) selfing occurs as an unavoidable by-product of pollinator movements within inflorescences (Harder & Barrett 1995; Johnston et al. 2009). In the more highly selfing mixed mating species (mode = 0.35), floral mechanisms may readily facilitate selfing to ensure reproduction in unpredictable pollination environments but some opportunity for pollinator visitation and outcrossing occurs.

**References**

Barrett, S.C.H. (2003). Mating strategies in flowering plants: the outcrossing-selfing paradigm and beyond. *Phil. Trans. R. Soc. Lond. B*, 358, 991—1004.

Barrett, S.C.H. & Harder, L.D. (1996). Ecology and evolution of plant mating. *Trends in Ecology &* *Evolution*, 11, 73-79.

Eckert, C.G., Samis, K.E., & Dart, S. (2006). Reproductive assurance and the evolution of uniparental reproduction in flowering plants. In: Ecology and evolution of flowers (ed. Harder, L.D. and Barrett, S.C.H.). Oxford University Press.

Goodwillie, C., Kalisz, S., & Eckert, C.G. (2005). The evolutionary enigma of mixed mating systems in plants: Occurrence, theoretical explanations, and empirical evidence. *Annu. Rev. Ecol. Evol. Syst.* 36, 47–79.

Goodwillie, C., Sargent, R.D., Eckert, C.G., Elle, E., Geber, M.A., Johnston, M.O., ... & Winn, A. A. (2010). Correlated evolution of mating system and floral display traits in flowering plants and its implications for the distribution of mating system variation. *New Phytologist*, 185, 311-321.

Igic, B., & Kohn, J.R. (2006). The distribution of plant mating systems: study bias against obligately outcrossing species. *Evolution*, 60, 1098—1103.

Johnston, M. O. 1998. Evolution of intermediate selfing rates in plants: pollination ecology versus deleterious mutations. *Genetica*, 103, 267-278.

Johnston, M.O., E. Porcher, P.-O. Cheptou, C.G. Eckert, E. Elle, M.A. Geber, S. Kalisz, J. K. Kelly, D.A. Moeller, M. Vallejo-Marín, and A.A. Winn. 2009. Correlations among fertility components can maintain mixed mating in plants. *American Naturalist*, 173, 1-11.

Lande, R., and Schemske, D.W. (1985). The evolution of self-fertilization and inbreeding depression in plants. 1. Genetic models. *Evolution,* 39, 24–40.

Porcher, E., and R. Lande. 2005. The evolution of self-fertilization and inbreeding depression under pollen discounting and pollen limitation. *J. Evol. Biol*., 18, 497–508.

Raduski, A. R., Haney, E. B., & Igić, B. (2012). The expression of self-incompatibility in angiosperms is bimodal. *Evolution*, 66, 1275-1283.

Schemske, D.W., & Lande, R. (1985). The evolution of self-fertilization and inbreeding depression in plants. 2. Empirical observations. *Evolution,* 39, 41–52.

Vallejo-Marín, M. & Uyenoyama, M.K. (2004) On the evolutionary costs of self-incompability: incomplete reproductive compensation due to pollen limitation. *Evolution*, 58, 1924–1935.

Vogler, D.W. & Kalisz, S. (2001). Sex among the flowers: the distribution of plant mating systems. *Evolution*, 55, 202-204.

**Table S1.1.** Results of normal mixture models with parameter estimates for the means (μ), standard deviations (σ), and probabilities (π) of each distribution. The probabilities (π) describe estimates of the fraction of observations belonging to each distribution. Likelihood ratio tests were used to determine whether the addition of normal distributions was significant compared to the next most simple model (e.g. 4 vs. 3 distributions). Likelihood ratio tests used a chi-square distribution with the number of degrees of freedom equal to the difference in the number of parameters between competing models.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Number of normal distributions | | | | |
|
| Parameter estimates | 1 | 2 | 3 | 4 | 5 |
| m1 | 0.666 | 0.477 | 0.045 | 0.039 | 0.037 |
| m2 |  | 0.906 | 0.623 | 0.346 | 0.382 |
| m3 |  |  | 0.923 | 0.675 | 0.61 |
| m4 |  |  |  | 0.919 | 0.711 |
| m5 |  |  |  |  | 0.911 |
| s1 | 0.309 | 0.288 | 0.043 | 0.037 | 0.036 |
| s2 |  | 0.079 | 0.221 | 0.131 | 0.153 |
| s3 |  |  | 0.062 | 0.088 | 0.033 |
| s4 |  |  |  | 0.072 | 0.042 |
| s5 |  |  |  |  | 0.076 |
| p1 | 1 | 0.558 | 0.11 | 0.103 | 0.1 |
| p2 |  | 0.442 | 0.533 | 0.17 | 0.208 |
| p3 |  |  | 0.357 | 0.266 | 0.079 |
| p4 |  |  |  | 0.461 | 0.117 |
| p5 |  |  |  |  | 0.496 |
|  |  |  |  |  |  |
| -2logL | 238.6 | 40.2 | -56.1 | -75.9 | -75 |
| P |  | < 0.0001 | < 0.0001 | < 0.001 | 0.825 |

**Figure S1.2.** Cumulative frequency distribution plots of outcrossing rate (tm) at multiple intervals during the history of research using molecular markers to infer mating system parameters: 1980, 1985 (from the landmark publication of Schemske & Lande 1985), 1990, 2000, and 2012. The two lower panels show histograms of the distributions for 1985 and 2012 for comparison.



**Appendix S2.** Ordinary least squares and phylogenetic regression models for outcrossing rate (tm) as the dependent variable and biogeographic (latitude and biome) and life form (life history and growth form) as independent variables. For the OLS analysis, we also show results from an analysis that included only angiosperms. The columns show adjusted R2 and the F statistic for the OLS analyses and the R2 and P value for the phylogenetic regression models.The rows show the four univariate models and the one multivariate analysis (including all four variables). Values in bold are significant at P< 0.05.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Ordinary Least Squares | | | | | | | | | Phylogenetic Regression | |
|  | All Species | | | Only Angiosperms | | | Only Animal-pollinated | | | All Species | |
| Model Factors | Adj. R2 | Df | F | Adj. R2 | Df | F | Adj. R2 | DF | F | Adj. R2 | P value |
| Latitude | **0.01** | **1, 429.8** | **6.48** | **0.03** | **1, 386.6** | **11.33** | **0.03** | **1, 345.5** | **10.70** | 0.00 | 0.5003 |
| Biome | **0.05** | **7, 356.9** | **3.99** | **0.05** | **7, 322.5** | **3.68** | **0.06** | **7, 286.4** | **3.48** | -0.01 | 0.8028 |
| Life History | **0.25** | **4, 484** | **41.51** | **0.23** | **4, 435** | **34.48** | **0.24** | **4, 391** | **32.96** | **0.30** | **<0.0001** |
| Growth Form | **0.24** | **4, 487** | **38.73** | **0.21** | **4, 438** | **30.2** | **0.19** | **4, 394** | **24.37** | **0.33** | **<0.0001** |
| Latitude + Biome + Life History + Growth Form | **0.37** | **16, 334.9** | **13.91** | **0.36** | **16, 302.9** | **12.29** | **0.34** | **16, 266.8** | **10.04** | **0.13** | **<0.0001** |

**Appendix S3.** Results of F tests from ordinary least squares regressions involving individual variables (univariate analyses) and the combined set of four variables (multivariate analysis). P values represent the mean from 1000 resampled datasets and the 95% confidence interval of those P values is also shown as a range. Results are presented for all species as well as for the analyses that include only angiosperms.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Ordinary Least Squares | | | | | | | | | | | | | | |
|  | All Species | | | | | Only Angiosperms | | | | | Only Animal-pollinated | | | | |
| Source | df | SS | F | P | P 95% CI | df | SS | F | P | P 95% CI | df | SS | F | P | P 95% CI |
| Univariate Analyses | | | | | | | | | | | | | | | |
| Latitude | 1.0 | 0.6 | 6.5 | 0.013 | 0.0039-0.0299 | 1.0 | 1.1 | 11.3 | 0.001 | 0.0002-0.0026 | 1.0 | 1.0 | 10.7 | 0.0014 | 0.0003-0.0038 |
| Residuals | 429.8 | 39.8 |  |  |  | 386.6 | 37.1 |  |  |  | 345.5 | 31.2 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Biome | 7.0 | 2.5 | 4.0 | 0.0006 | 0.00002-0.0028 | 7.0 | 2.5 | 3.7 | 0.0015 | 0.00006-0.0069 | 7.0 | 2.2 | 3.5 | 0.0021 | 0.0001-0.0076 |
| Residuals | 356.9 | 32.6 |  |  |  | 322.5 | 30.8 |  |  |  | 286.4 | 25.4 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Life History | 4.0 | 12.1 | 41.5 | <<0.0001 | <<0.0001 | 4.0 | 10.7 | 34.5 | <<0.0001 | <<0.0001 | 4.0 | 9.7 | 33.0 | <<0.0001 | <<0.0001 |
| Residuals | 484.0 | 35.4 |  |  |  | 435.0 | 33.9 |  |  |  | 391.0 | 28.7 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Growth Form | 4.0 | 11.5 | 38.7 | <<0.0001 | <<0.0001 | 4.0 | 9.7 | 30.2 | <<0.0001 | <<0.0001 | 4.0 | 7.7 | 24.4 | <<0.0001 | <<0.0001 |
| Residuals | 487.0 | 36.3 |  |  |  | 438.0 | 35.2 |  |  |  | 394.0 | 31.1 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Multivariate Analysis | | | | | | | | | | | | | | | |
| Biome | 7.0 | 0.0 | 1.1 | 0.3918 | 0.181-0.6408 | 7.0 | 0.5 | 1.1 | 0.4026 | 0.1829-0.6416 | 7.0 | 0.5 | 1.1 | 0.3714 | 0.1685-0.6200 |
| Life History | 4.0 | 3.4 | 14.1 | <<0.0001 | <<0.0001 | 4.0 | 3.4 | 13.4 | <<0.0001 | <<0.0001 | 4.0 | 3.2 | 12.9 | <<0.0001 | <<0.0001 |
| Growth Form | 4.0 | 2.5 | 10.3 | <<0.0001 | <<0.0001 | 4.0 | 2.3 | 9.1 | <<0.0001 | <<0.0001 | 4.0 | 1.4 | 5.7 | 0.0003 | 0.00004-0.0009 |
| Latitude | 1.0 | 0.5 | 0.1 | 0.7819 | 0.485-0.9934 | 1.0 | 0.0 | 0.1 | 0.7773 | 0.4617-0.9933 | 1.0 | 0.0 | 0.1 | 0.7597 | 0.4376-0.9893 |
| Residuals | 334.9 | 20.2 |  |  |  | 302.9 | 19.4 |  |  |  | 266.8 | 16.4 |  |  |  |

**Appendix S4.** Estimates for coefficients of outcrossing rate from ordinary least squares (OLS) models for each variable individually. The results presented for OLS models are the means of 1000 model runs after resampling the dataset. Phylogenetic regression was conducted using phylogenetic least squares (PGLS) regression for latitude; for categorical variables we performed phylogenetic independent contrasts using the brunch algorithm. The results presented for phylogenetic analyses are the means from 10 datasets. For the OLS results we show least square means and 95% confidence intervals. For phylogenetic analyses, we present the intercept (value of first factor) and deviations for subsequent factors.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | OLS | | | | | | | | | Phylogenetic Regression | |
|  | All Species | | | Only Angiosperms | | | Only Animal-pollinated | | | All Species | |
|  | Estimate | Upper CI | Lower CI | Estimate | Upper CI | Lower CI | Estimate | Upper CI | Lower CI | Estimate | P-value |
| Intercept | 0.77 | 0.844 | 0.697 | 0.786 | 0.863 | 0.709 | 0.786 | 0.864 | 0.709 | 0.557 | 0.05 |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Latitude | -0.003 | -0.005 | -0.001 | -0.004 | -0.002 | -0.006 | -0.010 | -0.002 | -0.019 | 0.001 | 0.758 |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Biome categories |  |  |  |  |  |  |  |  |  |  |  |
| Desert/arid scrub | 0.72 | 0.833 | 0.607 | 0.716 | 0.833 | 0.598 | 0.723 | 0.843 | 0.603 | -- | -- |
| Temperate grassland | 0.678 | 0.805 | 0.552 | 0.677 | 0.807 | 0.548 | 0.707 | 0.835 | 0.579 | 0.001 | 0.857 |
| Mediterranean/chapparal | 0.56 | 0.629 | 0.49 | 0.533 | 0.607 | 0.46 | 0.543 | 0.618 | 0.468 | -0.034 | 0.551 |
| Boreal forest/taiga | 0.819 | 0.946 | 0.692 | 0.699 | 1.127 | 0.271 | 0.670 | 1.009 | 0.332 | 0.031 | 0.283 |
| Tropical Savanna | 0.838 | 0.996 | 0.681 | 0.839 | 1 | 0.678 | 0.855 | 1.015 | 0.695 | -0.078 | 0.215 |
| Temperate deciduous forest | 0.664 | 0.722 | 0.606 | 0.654 | 0.714 | 0.593 | 0.646 | 0.712 | 0.579 | 0.059 | 0.334 |
| Tropical seasonal forest | 0.731 | 0.841 | 0.621 | 0.731 | 0.843 | 0.619 | 0.727 | 0.837 | 0.618 | -0.056 | 0.298 |
| Tropical rainforest | 0.764 | 0.834 | 0.694 | 0.76 | 0.834 | 0.686 | 0.754 | 0.827 | 0.681 | -0.04 | 0.562 |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Life History Categories |  |  |  |  |  |  |  |  |  |  |  |
| Annual | 0.402 | 0.45 | 0.354 | 0.402 | 0.451 | 0.352 | 0.409 | 0.459 | 0.358 | -- | -- |
| Biennial | 0.597 | 0.799 | 0.383 | 0.589 | 0.804 | 0.374 | 0.589 | 0.799 | 0.380 | 0.272 | 0 |
| Semelparous perennial | 0.759 | 0.959 | 0.559 | 0.759 | 0.966 | 0.552 | 0.758 | 0.960 | 0.557 | -0.078 | 0.463 |
| Iteroparous perennial | 0.761 | 0.789 | 0.732 | 0.746 | 0.778 | 0.714 | 0.746 | 0.778 | 0.713 | -0.073 | 0.599 |
| Variable | 0.396 | 0.633 | 0.159 | 0.396 | 0.64 | 0.151 | 0.194 | 0.502 | -0.114 | -- | -- |
| Appendix S4, continued |  |  |  |  |  |  |  |  |  |  |  |
|
| Growth Form Categories |  |  |  |  |  |  |  |  |  |  |  |
| Herbaceous | 0.529 | 0.564 | 0.495 | 0.529 | 0.565 | 0.493 | 0.540 | 0.577 | 0.502 | -- | -- |
| Vine | 0.476 | 0.624 | 0.327 | 0.476 | 0.631 | 0.322 | 0.476 | 0.630 | 0.323 | 0.176 | 0.003 |
| Shrub | 0.74 | 0.806 | 0.675 | 0.74 | 0.809 | 0.672 | 0.738 | 0.808 | 0.667 | 0.127 | 0.052 |
| Tree | 0.852 | 0.893 | 0.81 | 0.852 | 0.903 | 0.802 | 0.841 | 0.895 | 0.787 | -0.156 | 0.066 |
| Variable | 0.096 | 0.631 | -0.439 | 0.096 | 0.652 | -0.46 | 0.096 | 0.648 | -0.456 | -- | -- |

**Appendix S5.** Results of biogeographic analyses conducted within categories of life histories and growth forms for which sufficient data were available. In each case, we present univariate analyses, where each factor was analyzed separately, and bivariate analyses, where a multiple regression was used. Analyses were conducted on 1000 resampled datasets and therefore the mean denominator degrees of freedom and the mean and range of P-values is shown.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | df (num) | df (den) | F | P (mean) | P range |
| **Annual Life History** | | | | | |
| **Univariate** | | | | | |
| Latitude | 1 | 84.4 | 0.018 | 0.901 | (0.86-0.92) |
| Biome | 7 | 62.4 | 1.784 | 0.107 | (0.10-0.11) |
|  |  |  |  |  |  |
| **Bivariate** | | | | | |
| Latitude | 1 | 57.4 | 0.059 | 0.832 | (0.59-0.87) |
| Biome | 7 | 57.4 | 1.787 | 0.109 | (0.08-0.12) |
|  |  |  |  |  |  |
| **Perennial Life History** | | | | | |
| **Univariate** | | | | | |
| Latitude | 1 | 321.3 | 0.239 | 0.627 | (0.56-0.69) |
| Biome | 7 | 267 | 0.495 | 0.836 | (0.79-0.87) |
|  |  |  |  |  |  |
| **Bivariate** | | | | | |
| Latitude | 1 | 263 | 1.019 | 0.316 | (0.30-0.36) |
| Biome | 7 | 263 | 0.493 | 0.839 | (0.81-0.85) |
|  |  |  |  |  |  |
| **Herbaceous Growth Form** | | | | | |
| **Univariate** | | | | | |
| Latitude | 1 | 193.9 | 2.548 | 0.119 | (0.08-0.19) |
| Biome | 7 | 146.5 | 2.232 | **0.048** | (0.03-0.10) |
|  |  |  |  |  |  |
| **Bivariate** | | | | | |
| Latitude | 1 | 140.9 | 0.122 | 0.749 | (0.61-0.93) |
| Biome | 7 | 140.9 | 1.971 | 0.088 | (0.01-0.20) |
|  |  |  |  |  |  |
| **Tree Growth Form** | | | | | |
| **Univariate** | | | | | |
| Latitude | 1 | 157.7 | 0.128 | 0.770 | (0.05-0.94) |
| Biome | 6 | 138.6 | 0.47 | 0.825 | (0.71-0.94) |
|  |  |  |  |  |  |
| **Bivariate** | | | | | |
| Latitude | 1 | 134.3 | 0.754 | 0.411 | (0.24-0.60) |
| Biome | 6 | 134.4 | 0.62 | 0.713 | (0.54-0.83) |

**Appendix S6.** Logistic regression and phylogenetic logistic regression models for self-incompatibility (versus self-compatibility) as the dependent variable and biogeographic (latitude and biome) and life form (life history and growth form) as independent variables. For the logistic regressions, we also show results from an analysis that included only angiosperms. The columns show Akaike’s Information Criterion (AIC) and the residual deviance for the logistic regressions and the AIC and alpha value for the phylogenetic logistic regression models.The rows show the four univariate models and the one multivariate analysis (including all four variables). Values in bold are significant at P<0.05.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Logistic Regression | | | | Phylogenetic Logistic Regression | |  |
|  | All Species | | Only Angiosperms | | All Species | |  |
| Model factors | AIC | Residual Deviance | AIC | Residual Deviance | AIC | alpha |  |
| Latitude | 339.07 | 335.07 | 333.48 | 329.48 | 302.05 | 0.61 |  |
| Biome | 308.96 | 292.96 | 304.73 | 288.73 | 281.71 | 0.52 |  |
| Life History | 362.16 | 352.16 | 354.67 | 344.67 | **279.20** | **0.44** |  |
| Growth Form | 374.97 | 364.97 | 364.78 | 354.78 | **294.57** | **0.15** |  |
| Latitude + Biome + Life History + Growth Form | 295.95 | 261.95 | 289.36 | 255.36 | 305.10 | 0.11 |  |
|  |  |  |  |  |  |  |  |

**Appendix S7.** Results of logistic regression analyses for the probability of self-incompatibility. For tests of deviance, we show the chi-square values, p values, and the 95% confidence interval of P values from the 1000 resampled datasets.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | All species | | | | Only Angiosperms | | | |  |
|  | df | χ2 | P | P 95% CI | df | χ2 | P | P 95% CI |  |
| Univariate Analyses | | | | | | | | |  |
| Latitude | 1 | 1.79 | 0.195 | 0.089-0.344 | 1 | 1.54 | 0.230 | 0.103-0.414 |  |
|  |  |  |  |  |  |  |  |  |  |
| Biome | 7 | 11.10 | 0.155 | 0.036-0.352 | 7 | 8.65 | 0.305 | 0.082-0.604 |  |
|  |  |  |  |  |  |  |  |  |  |
| Life History | 4 | 21.81 | <0.001 | 0.0001-0.001 | 4 | 23.85 | <0.001 | <0.0001-0.0002 |  |
|  |  |  |  |  |  |  |  |  |  |
| Growth Form | 4 | 11.75 | 0.021 | 0.009-0.039 | 4 | 16.44 | 0.003 | 0.001-0.006 |  |
|  |  |  |  |  |  |  |  |  |  |
| Multivariate Analyses | | | | | | | | |  |
| Latitude | 1 | 1.34 | 0.154 | 0.045-0.331 | 1 | 1.61 | 0.231 | 0.080-0.444 |  |
| Biome | 7 | 14.29 | 0.059 | 0.009-0.171 | 4 | 10.85 | 0.175 | 0.032-0.439 |  |
| Life History | 4 | 11.80 | 0.021 | 0.007-0.046 | 4 | 12.19 | 0.018 | 0.006-0.042 |  |
| Growth Form | 4 | 5.72 | 0.228 | 0.128-0.357 | 7 | 7.06 | 0.139 | 0.071-0.236 |  |
|  |  |  |  |  |  |  |  |  |  |

**Appendix S8.** Estimates of coefficients (and standard errors) from logistic regressions of the probability of self-compatibility for each of the four predictor variables. Results are shown for the models including all species as well as for models that included angiosperms only.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Logistic Regression | | | | | |  |
|  | All species | | | Only Angiosperms | | |  |
|  |  | Standard Errors | |  | Standard Errors | |  |
|  | Estimate | Upper SE | Lower SE | Estimate | Upper SE | Lower SE |  |
| Intercept | 0.63 | 0.77 | 0.47 | 0.63 | 0.77 | 0.47 |  |
|  |  |  |  |  |  |  |  |
| Latitude | 0.50 | 0.58 | 0.50 | 0.50 | 0.51 | 0.50 |  |
|  |  |  |  |  |  |  |  |
| Biome Categories |  |  |  |  |  |  |  |
| Desert/arid scrub | 0.85 | 0.96 | 0.63 | 0.86 | 0.96 | 0.63 |  |
| Temperate grassland | 0.76 | 0.91 | 0.54 | 0.76 | 0.91 | 0.54 |  |
| Mediterranean/chapparal | 0.66 | 0.78 | 0.54 | 0.65 | 0.76 | 0.51 |  |
| Boreal forest/taiga | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 |  |
| Tropical Savannah | 0.39 | 0.72 | 0.12 | 0.39 | 0.72 | 0.12 |  |
| Temperate deciduous forest | 0.69 | 0.78 | 0.58 | 0.69 | 0.79 | 0.58 |  |
| Tropical seasonal forest | 0.73 | 0.88 | 0.50 | 0.73 | 0.88 | 0.50 |  |
| Tropical rainforest | 0.69 | 0.81 | 0.55 | 0.67 | 0.80 | 0.53 |  |
|  |  |  |  |  |  |  |  |
| Life History Categories |  |  |  |  |  |  |  |
| Annual | 0.89 | 0.94 | 0.82 | 0.89 | 0.94 | 0.82 |  |
| Biennial | 0.78 | 0.97 | 0.37 | 0.78 | 0.97 | 0.37 |  |
| Semelparous perennial | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 |  |
| Iteroparous perennial | 0.67 | 0.73 | 0.61 | 0.66 | 0.72 | 0.59 |  |
| Variable | 0.75 | 0.98 | 0.25 | 0.75 | 0.98 | 0.25 |  |
|  |  |  |  |  |  |  |  |
| Growth Form Categories |  |  |  |  |  |  |  |
| Herbaceous | 0.79 | 0.84 | 0.73 | 0.79 | 0.84 | 0.73 |  |
| Vine | 0.72 | 0.92 | 0.41 | 0.72 | 0.92 | 0.41 |  |
| Shrub | 0.80 | 0.90 | 0.67 | 0.80 | 0.90 | 0.67 |  |
| Tree | 0.60 | 0.70 | 0.49 | 0.54 | 0.66 | 0.43 |  |
| Variable | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 |  |
|  |  |  |  |  |  |  |  |

**Appendix S9:** Mulitnomial logistic regression of life history and growth form on latitude

**Methods**

In order to test for latitudinal variation to life history (annual, biennial, semelparous perennial, iteroparous perennial, and varies) and growth form (herb, vine, shrub, tree, and varies), we used multinomial logistic regressions. To do this, we used the function ‘multinom’ in the ‘nnet’ package v. 7.3-8 in R (Venables & Ripley 2002). In both models, either life history or growth form was the dependent variable and latitude was the independent variable. Due to the potential for curvilinear relationships, we also included a variable for the squared value of latitude in a second set of models. Log-likelihood tests indicated that the inclusion of both latitude and latitude-squared provided the best fit for both life history and growth form (Life history: χ2 = 7.6, df = 1, *P* = 0.006; Growth form: χ2 = 30.8, df = 1, *P* <0.0001).

Because coefficients in multinomial models are dependent on a pre-defined base level (one of the life history or growth form categories), we determined coefficients for all combinations of variables within life history and growth form by releveling the model for each category, rerunning the model and outputting the coefficients. We assessed the significance of coefficients using two-tailed Z-tests.

**Results**

Life history and growth form both varied significantly with latitude. For life history, there was a linear increase in the likelihood of annual life histories with latitude (Table S9.1). This relationship also exhibited significant curvilinearity where annuals were most frequent at temperate latitudes (35-45°) (Figures S9.3 and S9.4; Table S9.1). For growth form, there was a linear increase in the likelihood of herbaceous growth forms with increase in latitude (Table S9.2). There was also a curvilinear relationship due largely in part to a marked increase in shrubs at subtropical to temperate latitudes (15-40°) (Figures S9.3 and S9.4; Table S9.2).

**References**

Venables, W. N. & B. D. Ripley. (2002). Modern applied statistics with S. Fourth Edition. Springer, New York

**Table S9.1.** Coefficients for multinomial logistic regression of life history and latitude plus the squared value of latitude. Coefficients are in units of log odds (logit). Columns represent the level used as the base level for the run of the model to generate the coefficient values. The coefficient to the left of the slash is for latitude and the coefficient to the right of the slash is for latitude-squared. Bolded values are significant at P<0.05; italicized values are not quite significant: 0.05<P<0.1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Annual | Biennial | Semelparous perennial | Iteroparous perennial | Varies |
| Annual |  | **0.2567 / -0.0034** | **-0.1649 /** 0.0009 | **0.1612 / -0.0019** | **0.3746 / -0.0056** |
| Biennial | **-0.2567 / 0.0034** |  | **-0.4216 / 0.0043** | *-0.0956 /* 0.0016 | 0.1179 / -0.0021 |
| Semelparous perennial | **0.1650**  / -0.0009 | **0.4216 / -0.0044** |  | **0.3261 / -0.0028** | **0.5395 / -0.0065** |
| Iteroparous perennial | **-0.1612 / 0.0018** | *0.0955*  / -0.0016 | **-0.3261 / 0.0028** |  | **0.2134 / -0.0037** |
| Variable | **-0.3746 / 0.0056** | -0.1178 / 0.0022 | **-0.5395 / 0.0065** | **-0.2135 / 0.0038** |  |

**Table S9.2.** Coefficients for multinomial logistic regression of growth form and latitude plus the squared value of latitude. Coefficients are in units of log odds (logit). Columns represent the level used as the base level for the run of the model and its output. The coefficient to the left of the slash is for latitude and the coefficient to the right of the slash is for latitude-squared. Bolded values are significant at P<0.05; italicized values are not quite significant: 0.05<P<0.1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Herb | Vine | Shrub | Tree | Varies |
| Herb |  | **-0.1487 / 0.0038** | **-0.221 / 0.0054** | **0.1613 / -0.0017** | **-6.251 / 0.2738** |
| Vine | **0.1784 / -0.0042** |  | -0.0534 / 0.0014 | **0.3272 / -0.0057** | **-6.09 / 0.2698** |
| Shrub | **0.2284 / -0.0055** | *0.0777* / -0.0018 |  | **0.387 / -0.0072** | **-6.026 / 0.2682** |
| Tree | **-0.1591 / 0.0016** | **-0.3093 / 0.0054** | **-0.3824 / 0.0071** |  | **-6.412 / 0.2754** |
| Variable | **6.826 / -0.2983** | **9.719 / -0.4461** | **8.417 / -0.3831** | **68.37 / -3.55** |  |

**Figure S9.3.** Distribution of life history strategies and growth forms in the database. Each bar represents a bin of five degrees latitude. Life history strategy is found in the left panel and growth form is found in the right panel. The bar is divided based on the proportion of the different categories of life history strategy or growth form. The width of the bar is proportional to the number of data points occurring in the latitudinal bin. The five categories of life history in the left panel are represented by the following colors: annual = green, biennial = orange; semelparous perennial = light blue, iteroparous perennial = gray, and varies = red. The five categories of growth form in the right panel are represented by the following colors: herb = dark blue, vine = green, shrub = pink, tree = gray, and varies = red.



**Figure S9.4**. Distribution of life history strategies and growth forms output from the multinomial logistic regression including both latitude and latitude squared. Each bar represents a prediction for the proportion of life history or growth form categories at each latitude. Life history strategy is found in the left panel and growth form is found in the right panel. The five categories of life history in the left panel are represented by the following colors: annual = green, biennial = orange; semelparous perennial = light blue, iteroparous perennial = gray, and varies = red. The five categories of growth form in the right panel are represented by the following colors: herb = dark blue, vine = green, shrub = pink, tree = gray, and varies = red.



**Appendix S10.** Tests of phylogenetic signal in latitude and outcrossing rate using Blomberg’s K. The upper half of the table shows analyses of all species (both angiosperms and gymnosperms) for each of 10 resampled datasets; the lower half shows calculations for angiosperms only. Values of K quantify the degree of phylogenetic signal and Z-scores determine the difference between observed and expected variances for calculations of independent contrasts, along with the p-value for those comparisons.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| All species | | | | | | |
|  | Latitude | | | Outcrossing rate | | |
|  | K | Z-score | p-value | K | Z-score | p-value |
| Dataset 1 | 0.199 | -12.73 | 0.001 | 0.153 | -8.84 | 0.001 |
| Dataset 2 | 0.094 | -3.77 | 0.001 | 0.102 | -4.75 | 0.001 |
| Dataset 3 | 0.136 | -9.40 | 0.001 | 0.140 | -8.54 | 0.001 |
| Dataset 4 | 0.111 | -5.02 | 0.001 | 0.100 | -4.94 | 0.001 |
| Dataset 5 | 0.096 | -3.84 | 0.001 | 0.109 | -4.39 | 0.001 |
| Dataset 6 | 0.101 | -4.25 | 0.001 | 0.105 | -4.18 | 0.001 |
| Dataset 7 | 0.106 | -4.43 | 0.001 | 0.103 | -4.51 | 0.001 |
| Dataset 8 | 0.107 | -6.06 | 0.001 | 0.109 | -5.45 | 0.001 |
| Dataset 9 | 0.110 | -6.34 | 0.001 | 0.127 | -6.67 | 0.001 |
| Dataset 10 | 0.136 | -8.89 | 0.001 | 0.140 | -8.59 | 0.001 |
| Average | 0.119 | -6.47 | 0.001 | 0.119 | -6.09 | 0.001 |
|  | | | | | | |
| Only Angiosperms | | | | | | |
|  | Latitude | | | Outcrossing rate | | |
|  | K | Z-score | p-value | K | Z-score | p-value |
| Dataset 1 | 0.295 | -12.75 | 0.001 | 0.145 | -7.12 | 0.001 |
| Dataset 2 | 0.117 | -3.78 | 0.002 | 0.132 | -3.90 | 0.001 |
| Dataset 3 | 0.270 | -12.37 | 0.001 | 0.139 | -7.05 | 0.001 |
| Dataset 4 | 0.123 | -5.93 | 0.001 | 0.132 | -5.93 | 0.001 |
| Dataset 5 | 0.184 | -6.02 | 0.001 | 0.166 | -5.36 | 0.001 |
| Dataset 6 | 0.174 | -5.77 | 0.001 | 0.189 | -6.20 | 0.001 |
| Dataset 7 | 0.111 | -6.13 | 0.001 | 0.125 | -5.21 | 0.001 |
| Dataset 8 | 0.138 | -6.48 | 0.001 | 0.111 | -5.45 | 0.001 |
| Dataset 9 | 0.108 | -5.62 | 0.001 | 0.122 | -4.71 | 0.001 |
| Dataset 10 | 0.201 | -8.99 | 0.001 | 0.157 | -7.31 | 0.001 |
| Average | 0.172 | -7.39 | 0.001 | 0.142 | -5.82 | 0.001 |

**Appendix S11.** Results of phylogenetic logistic regressions for the probability of self-compatibility. We show the mean from 10 resampled datasets of estimates for each factor, standard errors (SE), Z values, and P values. We also show the range of values from the 10 resampled datasets. Results are shown for univariate analyses and for the multivariate analysis including all four predictors.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Estimate | Est. Range | SE | SE Range | Z | Z Range | P | P range |  |
| Univariate Analyses | | | | | | | | |  |
| Intercept | 0.57 | 0.08 - 1.73 | 0.44 | 0.36 - 0.59 | 1.23 | 0.21 - 2.90 | 0.389 | <0.001 - 0.833 |  |
| Latitude | 0.02 | -0.02 - 0.08 | 0.02 | 0.01 - 0.03 | 1.15 | -1.46 - 2.20 | 0.140 | 0.028 - 0.411 |  |
|  |  |  |  |  |  |  |  |  |  |
| Intercept | 1.48 | 1.31 - 1.61 | 0.42 | 0.41 - 0.43 | 3.5 | 3.22 - 3.76 | 0.001 | <0.001 - 0.001 |  |
| Biome | -0.12 | -0.14 - -0.09 | 0.07 | 0.07 - 0.07 | -1.7 | -1.98 - -1.54 | 0.096 | 0.048 - 0.202 |  |
|  |  |  |  |  |  |  |  |  |  |
| Intercept | 1.21 | 0.44 -1.53 | 0.52 | 0.43 - 0.54 | 2.34 | 0.82 - 3.04 | 0.085 | 0.002 - 0.411 |  |
| Life History | -0.45 | -0.61 - -0.10 | 0.13 | 0.12 - 0.15 | -3.43 | -4.43 - -0.80 | 0.061 | <0.001 - 0.426 |  |
|  |  |  |  |  |  |  |  |  |  |
| Intercept | 1.52 | 1.44 - 1.73 | 0.36 | 0.35 - 0.37 | 4.27 | 4.12 - 4.64 | <0.001 | <0.001 |  |
| Growth Form | -0.28 | -0.033 - -0.24 | 0.12 | 0.12 - 0.13 | -2.22 | -2.60 - -1.99 | 0.029 | 0.009 - 0.0047 |  |
|  |  |  |  |  |  |  |  |  |  |
| Multivariate analyses | | | | | | | | |  |
| Intercept | 1.72 | 1.37 - 1.96 | 0.84 | 0.79 - 0.94 | 2.04 | 1.66 - 2.28 | 0.044 | 0.023 - 0.096 |  |
| Latitude | 0 | -0.01 - 0.01 | 0.01 | 0.01 - 0.01 | -0.06 | -0.88 - 0.13 | 0.667 | 0.377 - 0.988 |  |
| Biome | -0.08 | -0.18 - -0.01 | 0.06 | 0.06 - 0.08 | -1.37 | -3.15 - -0.23 | 0.309 | 0.002 - 0.820 |  |
| Life History | -0.38 | -0.56 - -0.14 | 0.14 | 0.11 - 0.16 | -2.79 | -3.86 - -0.94 | 0.061 | <0.001 - 0.349 |  |
| Growth Form | -0.06 | -0.13 - 0.05 | 0.14 | 0.12 - 0.15 | -0.42 | -1.29 - 0.41 | 0.599 | 0.382 - 0.811 |  |
|  |  |  |  |  |  |  |  |  |  |