Variation in nuclear DNA content across environmental gradients: a quantile regression analysis

Abstract

The nuclear DNA content of angiosperms varies by several orders of magnitude. Previous studies suggest that variation in 2C DNA content (i.e. the amount of DNA in G1 phase nuclei, also referred to as the 2C-value) is correlated with environmental factors, but there are conflicting reports in the literature concerning the nature of these relationships. We examined variation in 2C DNA content for 401 species in the ecologically diverse California flora in relation to the mean July maximum temperature, January minimum temperature, and annual precipitation within the geographical ranges of these species. Species with small 2C-values predominate in all environments. Species with large 2C-values occur at intermediate July maximum temperatures, and decline in frequency at both extremes of the July temperature gradient, and with decreasing annual precipitation. Our analysis demonstrates the utility of quantile regression for statistical inference of complex distributions such as these. The method supports our observation that relationships between nuclear DNA content and environmental factors are stronger for species with large 2C-values.

Keywords

Genome size, evolution, macroecology, quantile regression, seed size, randomization tests.

INTRODUCTION

The 2C DNA content of angiosperms (i.e. the amount of DNA in G1 phase nuclei, also referred to as the 2C-value) varies by four orders of magnitude (Bennett *et al.* 1998). Several studies have examined variation in 2C DNA content across environmental gradients; however, these studies have often produced conflicting results (Table 1). For instance, early studies indicated a positive correlation with latitude (Bennett 1976; Levin & Funderburg 1979), which supported work by Avdulov (1931) and Stebbins (1971) who found that species with large chromosomes were more common at temperate latitudes (note that Levin & Funderburg (1979) also reported a positive correlation between chromosome volume and 2C DNA content). In contrast, Grime & Mowforth (1982) found a negative correlation between 2C DNA content and latitude for 169 species in the British flora, and Artic and Antarctic species examined by Bennett *et al.* (1982) were within the lower range of what is characteristic for temperate latitudes. Other studies have found no relationship (Teoh & Rees 1976; Sims & Price 1985; Creber *et al.* 1994). Positive, negative and nonsignificant relationships with altitude have also been reported (Table 1). Interestingly, Rayburn (1990) found that for *Zea mays* maximum 2C-values occurred at intermediate elevations with a trend toward reduced 2C-values at both high and low elevations. A synthesis of previous studies (Table 1) suggests that species with greater 2C DNA content may be more frequent in the temperate mid-latitudes and at intermediate elevations, with a trend towards species with reduced 2C-values at the equator and the poles and at high and low elevations (Bennett *et al.* 1982; Bennett 1987; Rayburn 1990).

Environmental factors that covary with altitude or latitude may represent selection pressures on nuclear DNA content (if such selection pressures do exist), yet they have been incorporated in only a few studies. For example, Wakamiya *et al.* (1993) reported a positive correlation between haploid *C*-values and local measurements of annual precipitation and a negative correlation between mean maximum temperature and *C*-value; Price *et al.* (1981)

Table 1 Summary of previous studies on the relationship between genome size and latitude, altitude, temperature, precipitation and seed size both at the population (pop.) and species (sp.) level. Correlations are either +, −, not significant (NS), or nonlinear (/\).

Variable	Cor.	Level	Description	Authors
Elevation	−	pop.	8 *Dactylis glomerata* in Spain	Creber *et al.* (1994)
Elevation	−	pop.	17 *Dactylis glomerata* in France and Italy	Reeves *et al.* (1998)
Elevation	−	pop.	8 *Zea mays* in New Mexico	Rayburn (1990)
Elevation	+	pop.	12 *Zea mays* South-western USA	Rayburn & Auger (1990)
Elevation	+	pop.	15 *Dasypyrum villosum*	Caceres *et al.* (1998)
Elevation	/\	pop.	23 *Zea mays* South-western USA	Rayburn (1990)
Elevation	NS	sp.	51 neotropical *Lonchocarpus* trees	Palomino & Sousa (2000)
Latitude	+	pop.	Several *Picea sitchensis*	Miksche (1967, 1971)
Latitude	−	pop.	Several *Festuca arundinacea*	Ceccarelli *et al.* (1992)
Latitude	−	pop.	North American cultivars of *Zea mays*	Rayburn *et al.* (1985)
Latitude	NS	pop.	Several *Picea glauca*	Teoh & Rees (1976)
Latitude	NS	pop.	10 *Dactylis glomerata*	Creber *et al.* (1994)
Latitude	+	sp.	Tropical vs. temperate grasses	Avdulov (1931)
Latitude	+	sp.	329 tropical vs. 527 temperate plants	Levin & Funderburg (1979)
Latitude	+	sp.	17 Poaceae and 15 Fabaceae crops	Bennett (1976)
Latitude	−	sp.	169 British plants	Grime & Mowforth (1982)
Latitude	−	sp.	23 Arctic plants	Bennet *et al.* (1982)
Latitude	NS	sp.	19 *Helianthus*	Sims & Price (1985)
Precipitation	+	pop.	24 *Microseris douglasii*	Price *et al.* (1981)
Precipitation	+	sp.	18 North American *Pinus*	Wakamiya *et al.* (1993)
Precipitation	NS	sp.	19 *Helianthus*	Sims & Price (1985)
Temperature	−	sp.	18 North American pines	Wakamiya *et al.* (1993)
Seed size	+	pop.	15 *Dasypyrum villosum*	Caceres *et al.* (1998)
Seed size	+	pop.	12 Soybean strains	Chung *et al.* (1998)
Seed size	+	sp.	121 British plants	Thompson (1990)
Seed Size	+	sp.	43 British plants	Grime *et al.* (1997)
Seed Size	+	sp.	22 *Crepis*	Jones & Brown (1976)
Seed Size	+	sp.	12 *Allium* and 6 *Vicia*	Bennett (1972)
Seed Size	+	sp.	18 North American pines	Wakamiya *et al.* (1993)
Seed Size	+	sp.	Several Poaceae and Fabaceae	Mowforth (1985)

reported a positive correlation between annual precipitation and 2C DNA content; and MacGillivray & Grime (1995) found that frost-resistant species tended to have greater 2C DNA contents. However, as discussed above, the relationship between nuclear DNA content and environmental factors may not be a linear function. Considering that species with small 2C-values are more common than species with large 2C-values (Bennett *et al.* 1998), median 2C DNA content may vary little across environmental gradients, while there may be a significant trend for species with the largest 2C DNA contents. The mechanism(s) underlying 2C DNA content by environment correlations are still speculative, but it should be noted that environmental factors might impose selection pressures on physiological and life history traits that are correlated with 2C DNA content (Bennett 1976, 1987; Grime & Mowforth 1982; Grime *et al.* 1985, 1997; Mowforth & Grime 1989; MacGillivray & Grime 1995).

In this paper we examine variation in nuclear 2C DNA content with respect to July maximum temperature, January minimum temperature and the mean annual precipitation inside the ranges of 401 species in the California flora. The 2C DNA contents of the species selected vary over 100-fold; from 0.294 Gbp (for *Epilobium palustre,* about the same size as *Arabidopsis thaliana*) to 45.67 Gbp (*Smilacina racemosa*). A synthesis of previous studies (Table 1) led us to the hypothesis that species with large nuclear DNA contents may be rare in stressful environments, e.g. at both temperature extremes and with increasing aridity. We test this hypothesis using traditional least squares regression and a series of quantile regression functions (Cade *et al.* 1999; Koenker & Hallock 2001). The familiar least squares regression finds the best fit line for the mean of the dependent variable as a function of the independent variable; in contrast, quantile regression predicts the conditional quantile (percentile) of a dependent variable with respect to an independent variable. The 95th quantile regression can be thought of as a best-fit line that minimizes absolute residual errors, with the condition that 95% of the

observations fall below the line and 5% of the observations fall above the line. Significant differences for quantile regression slopes calculated for lower and upper quantiles indicate that the lower and upper bounds of the distribution are not parallel (Scharf *et al.* 1998). The same theory applies for polynomial quantile regression, and we demonstrate the utility of this technique here. We also used conventional statistical techniques to examine the relationship between variation in 2C DNA content and seed size, between species with annual and perennial growth habits, and between diploid and polyploid species.

METHODS

Genome size

We obtained estimates of 4C DNA content for 401 species in the California flora from the Angiosperm DNA *C*-value database (Bennett *et al.* 2001). We compare the DNA content of G_1 phase nuclei, which is also referred to as the 2C DNA content or 2C-value. 2C-values were obtained by halving the 4C DNA values presented in the database (the 4C-DNA value is the amount of DNA in replicated nuclei that have not yet divided). We chose to use 2C DNA amounts instead of haploid genome sizes because they represent the actual DNA content of the nucleus, regardless of ploidy levels. Moreover, the haploid genome size cannot be calculated for any species whose ploidy level is unknown.

2C DNA contents are frequently estimated by flow cytometry of isolated nuclei with DNA binding fluorochromes (Johnston *et al.* 1999; Bennett *et al.* 2000). A frequency distribution of nuclear fluorescence intensities for a random sample of isolated nuclei from the same individual is bi-modal, with a larger peak with lower fluorescence intensity equivalent to the G_1 nuclear genomes (2C), and a smaller peak with greater fluorescence intensity, which represents replicated nuclear genomes that have not yet divided (4C). The pg of DNA per nucleus in each peak is estimated by comparing the unknown to a known reference standard. We report 2C DNA contents in gigabase pairs (Gbp, 1 Gb = 1000 megabases = 10^9 bases) using the conversion factor 1 pg = 0.980 Gbp suggested by Bennett *et al.* (2000) and supported by Cavalier-Smith (1985). Ploidy levels were also obtained from the Angiosperm *C*-value database. Baker (1972) collected seed size data for 2003 California plant species, and 148 of these species were among the 401 species in the California flora with estimates of 2C DNA content. Classification of annual or perennial life history followed Hickman (1993). The status of each species as native or exotic followed the CalFlora database (CalFlora 2001). A list of species used in this study, along with their 2C DNA amount, mean environmental parameters, seed sizes, ploidy and classification as annuals or perennials and as native or exotic is available at www.stanford.edu/~dackerly/archive/calgenome.html.

Geographical information system (GIS) analysis

We used a GIS to estimate the mean July maximum temperature, January minimum temperature and annual precipitation within the geographical range of each species. Information on the presence or absence of each species inside 35 different floristic subregions of California combined with their upper and lower elevation limits (as reported in the Jepson Manual; Hickman 1993), were used to create coarse-scale species range maps using a digital elevation model of California (United States Geological Survey) and ArcView GIS software (ESRI, Redlands, California, USA). Climate maps were obtained from the Oregon State University PRISM project (Daly *et al.* 1994, 1997). The climate maps were created using local interpolation of mean climatic parameters between National Oceanographic and Atmospheric Administration (NOAA) weather stations. Mean July maximum temperature, mean January mimimum temperature and the mean annual precipitation inside the range of each species was calculated by intersecting the species range maps with the climate maps (Knight & Ackerly 2001). We used these estimates to quantify differences between species for realized climatic niche parameters (*sensu* Austin *et al.* 1990; Westman 1991; Franklin 1998).

Quantile regression

Quantile regression, as described by Koenker & Bassett (1978), extends classical least squares regression to an ensemble of models for conditional quantile functions. For least squares functions, squared residual errors are minimized with respect to the conditional mean. Quantile regression functions are estimated by minimizing an asymmetrically weighted sum of residual errors (Koenker & Bassett 1978; Cade *et al.* 1999; Koenker & Hallock 2001). Quantile regression is insensitive to heteroscedastic errors and dependent variable outliers (Koenker & Bassett 1978; Buchinsky 1991; Koenker & Portnoy 1996), both of which significantly affect least squares models (Cade & Richards 1996). For a random variable Y, the τth quantile is defined by a value y such that the probability of finding a smaller y is less than or equal to τ, and the probability of finding a larger y is less than or equal to $1 - \tau$. Similarly, the τth quantile regression function, $\mathrm{B}(\tau)$, corresponds to a linear or quadratic function fit through the data such that approximately τ proportion of the observations are less than $\mathrm{B}(\tau)$ and $1 - \tau$ proportion of the observations are greater than $\mathrm{B}(\tau)$. Estimates $b(\tau)$ of $\mathrm{B}(\tau)$ are obtained by minimizing the absolute values of residuals where positive

residuals are given weights equal to τ and negative residuals are given weights equal to $1 - \tau$ (see Koenker & Bassett 1978; Cade *et al.* 1999; Koenker & Hallock 2001 for mathematical details and a description of algorithms). Analogous to percentage notation, $100*\tau$ is often used to describe the τth quantile, and this convention will be followed here.

Examining quantile regression functions across the entire range of τ provides a more complete view of the relationship between x and y than what is captured by least squares or individual quantile regression functions. Econometricians have developed the method over the last 20 years and only recently has the technique been applied to ecological data (Koenker *et al.* 1994; Terrel *et al.* 1996; Scharf *et al.* 1998; Cade *et al.* 1999). As Cade *et al.* (1999) suggested, quantile regression is particularly useful in exploratory and inferential analyses concerning limiting factors. We used the "quantreg" library of the R statistical analysis software (distributed freely at http://www.R-project.org/) to perform this analysis.

Analyses were performed separately for each environmental parameter. July maximum temperature and January minimum temperature were included as quadratic terms in our least squares and quantile regressions. In either case a negative quadratic coefficient supports our hypothesis that (1) the relationship between 2C DNA content and temperature is nonlinear, and (2) the conditional mean 2C DNA content (in the case of least squares regression) or the quantile τ (in the case of quantile regression) decreases at both extremes of the temperature scale (if the maximum value of the model occurs within the range of the environmental variable). We estimated the complete series of quantile regression functions from the 5th to the 95th quantile. A similar analysis was performed for the relationship between mean annual precipitation and 2C DNA content, except in this case there was no *a priori* reason to expect that there would be a quadratic relationship; therefore, only the linear term was modelled. Confidence intervals were calculated with 1000 replicate bootstraps. Koenker & Hallock (2001) demonstrate that this method is conservative and not significantly different from the rank inversion test for calculating confidence intervals for quantile regression statistics.

To estimate the probability that our results were an artifact of sampling error, we created our own null distribution by recalculating regressions across 9999 random repairings of the x and y variables (a permutation or randomization test). All of the linear and quadratic coefficients were tested against the null model, but for the continuous quantile regressions, only the midpoint and extreme quantiles found to be significant from the bootstrap method were tested against the null model. Significance values (P values) were assigned by summing the number of times the test statistic (the quadratic coefficient for example) was greater (in the case of the quadratic coefficient, more negative) than the observed statistic and dividing by 10000 (9999 plus one for the observed relationship) as recommended by Manly (1997).

Previous investigators (Table 1) have found a positive correlation between the log of seed size and the log of 2C DNA content, so we transformed our data similarly prior to the seed size analysis, and used Pearson's correlation coefficient (Sokal & Rohlf 1995). We also performed the climate analyses with log transformed 2C-values, but the transformation did not change the results. Here we report analyses with the untransformed data. Differences in 2C DNA content between annuals and perennials, natives and exotics, and diploids and polyploids were analysed using Wilcoxon's nonparametric two-sample test (the Z statistic, Sokal & Rohlf 1995).

RESULTS

Temperature

The relationship between 2C DNA content and July maximum temperature is roughly a filled unimodal distribution (Fig. 1a). Species with small 2C-values predominate across the entire range of July maximum temperature, but species with large 2C-values tend to occupy ranges with July maximum temperatures near the midpoint of the scale. A standard quadratic polynomial regression fit to these data was significant ($F_{2,398} = 6.87$, $P = 0.001$). The quadratic coefficient was negative (coef. $= -0.0421$, $t_{2,398} = -2.88$, $P = 0.004$), indicating that mean 2C DNA content decreases as July maximum temperature increased or decreased from the 27°C vertex. When the model was run without the quadratic term, the linear coefficient was negative (coef. $= -0.22$), and the model was significant ($F_{1,399} = 5.33$, $P = 0.022$). Both the polynomial and the linear models were significantly different from regressions calculated for the randomized null model ($P < 0.001$ for each case).

The quadratic coefficient a for the 5th through 60th quantile regression was not significantly different from zero (Fig. 1b). Between the 5th and 50th quantiles the least squares estimate of a was significantly different, and more negative, than the quantile regression estimates. For quantile regressions between the 75th and 95th quantile, a became progressively more negative and significantly different from zero. This supports our hypothesis that environmental constraints on nuclear DNA content relate primarily to species with larger 2C DNA amounts. The 75th, 85th and 95th quadratic quantile regression coefficients were significantly different from quantile regression coefficients calculated for the randomized null model ($P \leq 0.001$, < 0.001

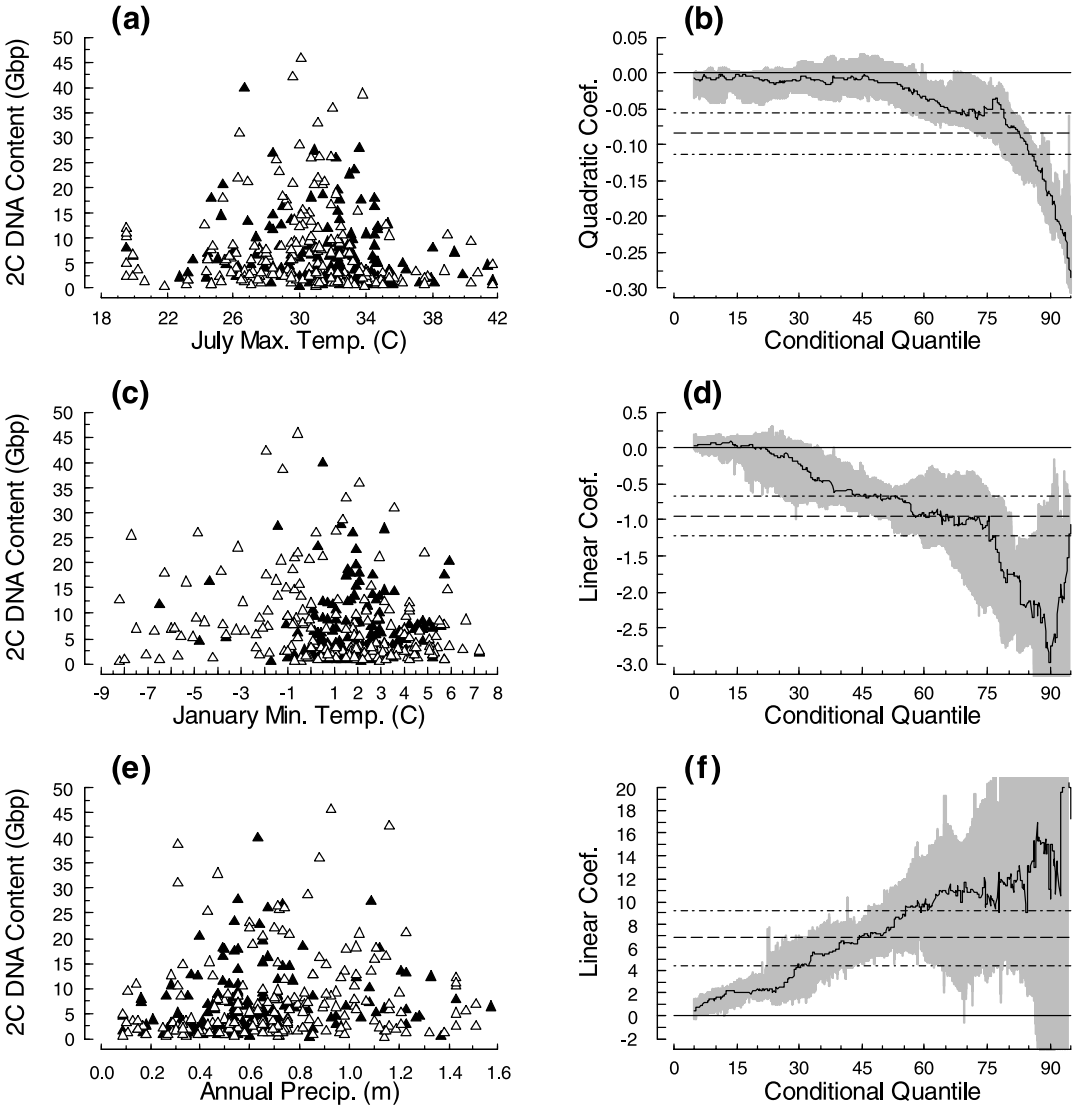

Figure 1 The relationship between 2CDNA content and July maximum temperature (a) January minimum temperature (c) and annual precipitation (e) 2CDNA contents are plotted as Gbp. Solid triangles are annual species and open triangles are perennials in (a), (c) and (e). The sensitivity of the quadratic coefficient for the July maximum temperature/2C DNA content analysis (b) and for the linear coefficient for the January minimum (d) and annual precipitation analysis (f) across the 5th to 95th quantiles of the data. Single dashed lines are the least squares estimates for the coefficients in (b), (d) and (f), the double dashed lines are the 95% confidence intervals for the least squares estimates. The grey area in (b), (d) and (f) represents the 95% confidence interval for the quantile regression estimates plotted as the black line.

and 0.015, respectively), as was the quadratic coefficient for the least squares model ($P < 0.001$).

July maximum temperature and January minimum temperature were not correlated ($F_{2,399} = 0.44$, $P =$ NS). This occurs because species growing on the coastal bluffs of northern California have the lowest July maximum temperature estimates but some of the warmest January minimum temperatures (Fig. 2a). There was a significant negative linear relationship of 2C DNA content vs. January minimum temperature (coef. $= -0.47$, $t_{2,399} = -3.923$, $P < 0.001$, Fig. 1c). Quantile regression analysis indicated that the trend for decreasing 2C DNA content with increasing January minimum temperature was stronger for species with larger 2C-values, and only significant above the 40th quantile (Fig. 1d). The linear coefficients for the least squares, and the 35th, 60th and 85th quantile regressions

were all significantly more negative than the coefficients calculated for the randomized null model ($P < 0.01$ in all cases).

Precipitation

There was a positive correlation between 2C DNA content and annual precipitation (least squares coef. = 3.42, $F_{1,399} = 8.26$, $P = 0.004$, Fig. 1e). Similar to the bivariate distribution for July maximum temperature, species with small 2C-values were common across the entire range of annual precipitation, while species with large 2C-values were infrequent at the two extremes. These results are not orthogonal to those for the temperature variables because there was a strong negative correlation between annual precipitation and July maximum temperature ($R = -0.64$, $N = 401$, $P < 0.001$, Fig. 2b) and January minimum temperature ($R = -0.38$, $N = 401$, $P < 0.001$). Linear coefficients for quantile regressions between the 5th and 85th quantile were positive, significantly different from zero, and increased with increasing quantiles (Fig. 1f). Quantile regressions above the 85th quantile were not significant. Quadratic coefficients for quantile regressions were only significant between the 88th and 94th quantiles. Linear coefficients calculated for the least squares and the 25th, 50th and 75th linear quantile regressions were all significantly different from coefficients calculated for the randomized null model ($P < 0.01$ in all cases).

Life history

There was a significant positive correlation between log transformed seed size and log transformed 2C DNA content ($R = 0.45$, $N = 148$, $P < 0.001$, Fig. 3). Forty of these species were also sampled by Thompson (1990). Our set of species overlaps, but our estimates of seed size are independent. There was no relationship between seed size and July maximum temperature, January minimum temperature or annual precipitation (results not shown).

One hundred and twenty-one of the species were native to the California flora. Exotic species had significantly lower 2C DNA contents ($Z_{280,121} = 2.69$, $P = 0.001$); however, the distributions of native and exotic species with respect to July maximum temperature, January minimum temperature and annual precipitation were similar (Figs 2a, b). There was, however, a disproportionate frequency of exotics in maritime climates characterized by low summer temperatures and relatively high January minimum temperatures.

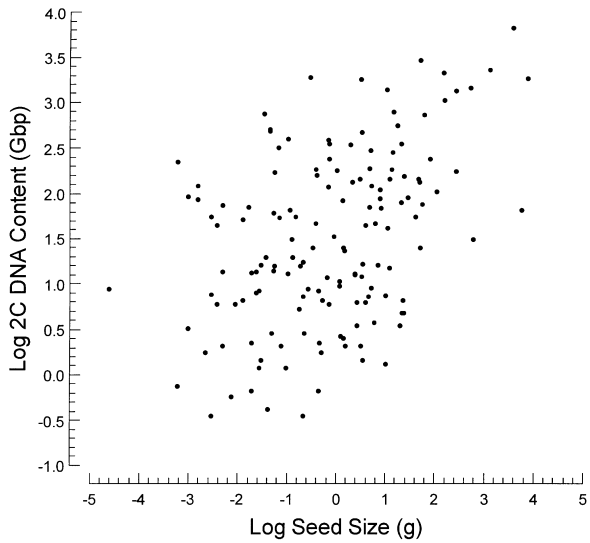

Figure 3 The relationship between log 2CDNA content (Gbp) and log seed size (g).

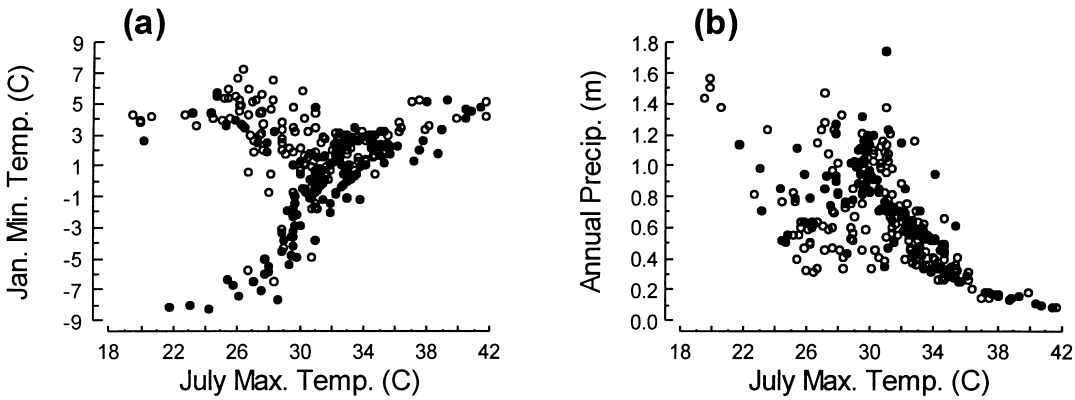

Figure 2 The relationship between July maximum temperature and January minimum temperature (a) and between July maximum temperature and annual precipitation (b) for the species in our study. Filled circles are native species and open circles are exotics.

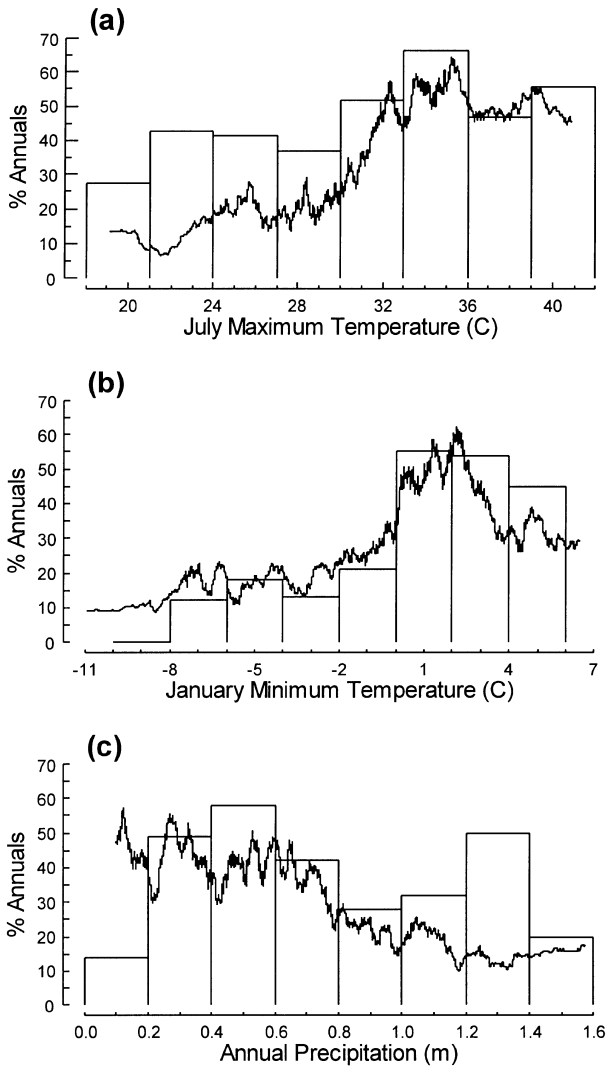

Figure 4 The percentage of annuals in our sample (columns) and the entire California flora (black line) as a function of species specific estimates for July maximum temperature (a) January minimum temperature (b) and annual precipitation (c).

When we calculated the least squares and quantile regressions for nuclear DNA content and environmental factors for natives and exotics separately, the results were similar in sign and magnitude to results for natives and exotics together (results not presented). Interestingly, the mean range size of native and exotic species was not significantly different (calculated as a part of our climate analysis, results not shown).

Fifty-one percent of the species sampled were perennials. The mean 2C DNA content of perennials (7.36 Gbp) was not significantly greater than that of annuals (6.19 Gbp, $Z_{199,195} = -0.91$, P = NS). Annuals typically had ranges with higher mean July maximum temperatures ($Z_{199,195} = 3.19$, $P = 0.001$), higher January minimum temperatures ($Z_{199,195} = 4.19$, $P < 0.001$) and lower annual precipitation ($Z_{199,195} = -2.15$, $P = 0.03$), which follows the pronounced trend for the entire California flora (Fig. 4a–c).

Of the 334 species of known ploidy, 215 were diploid and 119 were polyploid. Polyploid species had significantly greater 2C DNA contents compared with diploid species (mean 9.29 and 5.74 Gbp, respectively, $Z_{215,119} = -6.067$, $P < 0.001$). When analysed by Wilcoxon's two-sample test, the mean July maximum temperature inside the ranges of polyploid species was not significantly greater than that for diploid species ($Z_{215,119} = 0.50$, P = NS). However, polyploids did have significantly greater mean annual precipitation ($Z_{215,119} = -3.50$, $P < 0.001$) and lower January minimum temperatures estimates ($Z_{215,119} = 3.66$, $P < 0.001$).

DISCUSSION

Our analysis highlights the complexity of the relationships between nuclear DNA content and environmental conditions. Species with small 2C-values predominate in all environments. Species with large 2C-values tend to occur at intermediate July maximum temperatures, declining in frequency at both extremes along the July maximum temperature gradient, and also declining with decreasing annual precipitation. Our analysis required nontraditional statistical techniques not only because of the non-normality of the data but also because of the nature of our hypotheses (i.e. that species with large 2C-values would be infrequent in extreme environments). We demonstrate the utility of quantile regression in combination with randomization tests for statistical inference of complex data distributions and hypotheses such as these. The method supports our hypothesis that nuclear DNA content by environment relationships are stronger for species with larger 2C-values, with relatively little change in mean 2C-values. Our results indicate that shifts in 2C DNA content with environmental conditions depend on the range of environmental gradients sampled, which may help to explain why both positive and negative correlations have been found in the past.

If the analysis is performed with January minimum temperatures, species with large 2C-values become less frequent with increasing temperature, similar to the July maximum temperature analysis. Unlike the July maximum temperature analysis, the frequency of species with large 2C-values does not decrease at the coldest extreme. Species with low July maximum temperatures inside their geographical ranges do not necessarily have low January minimum temperatures; primarily as a result of the marine influence for species restricted to coastal areas (Fig. 2a).

This partially explains the discrepancy between the July maximum temperature and January minimum temperature analyses.

The combined effects of water stress and high temperature stress may amplify putative environmental effects on nuclear DNA content; and annual precipitation and July maximum temperature are strongly correlated at the high temperature extreme (Fig. 2b). Parallel to the July maximum temperature results, as annual precipitation decreased, the frequency of species with large 2C-values decreased. However, there was a small group of species with high mean annual precipitation estimates that also had small 2C-values. A large proportion of these species were from the coastal bluffs of northern California, a region that also has very cool July maximum temperatures. In addition, many of these species were exotics. Bennett *et al.* (1998) found that invasive weeds had smaller 2C DNA contents, and the same was true for our sample.

There may be significant allometric consequences related to variation in nuclear DNA content (Bennett 1976, 1987; Price 1988). This hypothesis primarily relies on strong positive correlations between nuclear DNA content, cell size and minimum generation time, and negative correlations with cell division rate and growth rate (Bennett 1971, 1972, 1977; Evans *et al.* 1972; Edwards & Endrizzi 1975; Van't Hof 1975; Bennett *et al.* 1983; Mowforth & Grime 1989). Nuclear DNA content tends to be greater in polypoid species, in species with large seeds, and in perennials compared with annuals (Bennett 1972, 1987; Jones & Brown 1976; Thompson 1990; Grime *et al.* 1997; Bennett *et al.* 1998; Chung *et al.* 1998). There is little doubt that these relationships are general and strong, but our mechanistic understanding for why they exist is still incomplete.

In the context of the life history and cellular correlations presented above, the decline in maximum 2C DNA values with increasing July maximum temperature or January minimum temperature, and with decreasing annual precipitation, might reflect an increased representation of diploids, annuals and/or small-seeded species. There is a strong trend for an increased abundance of annuals with increasing July maximum and January minimum temperatures, and decreasing annual precipitation, both for species in our analysis and across the entire California flora (Fig. 4), and previous investigators have reported that annual species have lower 2C-values than perennials (Bennett 1987; Bennett *et al.* 1998). However, in our sample, annual species did not have significantly lower 2C-values compared with perennials. Diploid species also had less 2C DNA and had significantly lower mean annual precipitation and January minimum temperature estimates, but there was not a significant difference for July maximum temperature. Variation in seed size was not correlated with July maximum temperature, January minimum temperature or annual precipitation. Therefore, the frequency of diploid species only partially contributes to the reduced abundance of species with large 2C-values in regions with high temperatures and low mean annual precipitation, while variation in seed size and the frequency of annuals does not contribute to the observed pattern.

Other studies have examined relationships between 2C DNA content and plant performance in response to artificial temperature treatments. For example, Thompson (1990) found a negative correlation between the minimum temperature for germination and 2C DNA content, which primarily reflected an absence of species with large 2C-values that required high temperatures for germination. In addition, MacGillivray & Grime (1995) found that species more resistant to artificial cold temperature treatments were more likely to have large 2C-values. These observations are consistent with our January minimum temperature analysis. MacGillivray & Grime (1995) do not propose a mechanism, but state that this relationship has arisen from "parallel effects of natural selection rather than any inherent functional relationships between genome size and frost sensitivity". It should be noted that selection pressures acting on nuclear DNA content alone and through functional relationships between nuclear DNA content and other aspects of cellular physiology may both contribute to the observed relationships with environmental factors.

Over two-thirds of the species sampled in our study were not native to California. Current estimates suggest that approximately one-eighth of the California flora is now composed of invasive or naturalized species (CalFora 2001). The enrichment of non-native species in our analyses occurs because there are more estimates of 2C DNA content for weedy European species. We felt that there was no *a priori* reason to exclude non-native species from our analysis. Although they may have had less time to reach their optimal distribution with respect to environmental parameters, this should not influence our results unless there is a bias for non-native species to live at one extreme of their tolerance breadth. Exotic species in our sample did have significantly lower 2C DNA contents, which concur with results of Bennett *et al.* (1998), but analyses of 2C DNA content and climate involving natives and exotics alone produced similar results. Interestingly, the average range size of native and non-native species was not significantly different.

Ecologists have given considerable attention to the problem of finding and interpreting boundaries, or limits, to complex bivariate distributions. There are statistical properties of the distribution of points within the "cloud" that may also be of ecological interest beyond what is captured by traditional mean regression techniques.

Quantile regression has been used for this endeavour, yet so far only for extreme quantiles of linear functions (Scharf *et al.* 1998; Cade *et al.* 1999). Estimating the entire ensemble of quantile regression functions ($5 < \tau < 95$) yields a more complete view of the relationship between two variables not encapsulated by changes in mean regression functions, and not necessarily observable by examining only the extreme quantile functions or by estimating limits by other means. The method was particularly useful in our analysis for providing quantitative evidence that trends towards reduced nuclear DNA content with increasing July maximum and January minimum temperature, and with decreasing annual precipitation, were stronger for species with larger 2C-values. The theory for quantile regression is well developed and is particularly appropriate for complex ecological data.

There should be a continued effort to obtain estimates of nuclear DNA contents for the native flora of North America in addition to exotics. As information on the climate and geographical distribution of plant species becomes more refined, analyses such as the ones presented here should be revisited. Environmental correlations with nuclear DNA content may represent functional relationships between cellular and whole plant physiology and environmental factors. However, these relationships may also arise from correlated selection pressures acting directly on nuclear DNA content. Regardless of their causal nature, variation in nuclear DNA content appears to be an important heritable trait that contributes to physiological, life history and climatic differentiation between species. In an age of functional genomics it deserves further attention.

ACKNOWLEDGEMENTS

We thank Roger Koenker for help with the quantile regression analysis; Dmitri Petrov, Elizabeth Hadly, Virginia Walbot and Johann Grielhuber for helpful discussion; Radika Bhaskar, Reto Nyffeler, Katherine Preston, Nathan Sanders, Dylan Schwilk and Ted Wong for helpful comments that improved the manuscript. This study was funded in part from a Tri-Agency (DOE, NSF, USDA) Training Grant in Plant Biology, an NSF Dissertation Improvement Grant (IBN-9902295; CAK) and fellowship support from the Center for Evolutionary Studies at Stanford University (CAK).

REFERENCES

Austin, M.P., Nicholls, A.O. & Margules, C.R. (1990). Measurement of the realized qualitative niche: environmental niches of five Eucalyptus species. *Ecol. Monographs*, 60, 161–177.

Avdulov, N.P. (1931). Karyo-systematishce untersuchungen der familie Gramineen. *Bull. Appl. Bot. Genet Plant Breed*, 44, 1–428.

Baker, H.G. (1972). Seed weight in relation to environmental conditions in California. *Ecology*, 53, 997–1010.

Bennett, M.D. (1971). The duration of meiosis. *Proc. Royal Soc. London B*, 178, 259–275.

Bennett, M.D. (1972). Nuclear DNA content and minimum generation time. *Proc. Royal Soc. London, Biol. Sci.*, 181, 109–135.

Bennett, M.D. (1976). DNA amount, latitude, and crop plant distribution. *Environ. Exp Bot.*, 16, 93–108.

Bennett, M.D. (1977). The time and duration of meiosis. *Phil. Trans. Royal Soc. London B*, 277, 201–226.

Bennett, M.D. (1987). Variation in genomic form in plants and its ecological implications. *New Phytologist*, 106, 177–200.

Bennett, M.D., Smith, J.B. & Lewis Smith, R.I. (1982). DNA amounts of angiosperms from the Antartic and South Georgia. *Environ. Exp. Bot.*, 22, 307–318.

Bennett, M.D., Heslop-Harrison, J.S., Smith, J.B. & Ward, J.P. (1983). DNA density in mitotic and meiotic metaphase chromosomes of plants and animals. *J. Cell Sci.*, 63, 173–179.

Bennett, M.D., Leitch, I.J. & Hanson, L. (1998). DNA amounts in two samples of angiosperm weeds. *Ann. Bot.*, 82, 121–134.

Bennett, M.D., Bhandol, P. & Leitch, I.J. (2000). Nuclear DNA amounts in angiosperms and their modern uses – 807 new estimates. *Ann. Bot.*, 86, 859–909.

Bennett, M.D., Cox, A.V. & Leitch, I.J. (2001). Angiosperm DNA. C-values database. [WWW document]. URL http://www.rbgkew.org.uk/cval/database1.html.Consulted:9/1/2001.

Buchinsky, M. (1991). The theory and practice of quantile regression. Dissertation. Harvard University, Cambridge, Massachusetts, USA.

Caceres, M.E., Pace, C.D., Mugnozza, G.T.S., Kotsonis, P., Ceccarelli, M. & Cionini, P.G. (1998). Genome size variations within *Dasypyrum villosum*: correlations with chromosomal traits, environmental factors and plant phenotypic characteristics and behavior in reproduction. *Theoret. Appl. Genet.*, 96, 559–567.

Cade, B.S. & Richards, J.D. (1996). Permutation test for least absolute deviation regression. *Biometrics*, 52, 886–902.

Cade, B.S., Terrell, J.W. & Schroeder, R.L. (1999). Estimating effects of limiting factors with regression quantiles. *Ecology*, 80, 311–323.

CalFlora (2001). CalFlora: Information on California plants for education, research and conservation. [web application]. (2000) Berkeley, California: The CalFlora Database [WWW document]. URL http://www.Calflora.OrgConsulted:12/15/2001.

Cavalier-Smith, T. (1985). *The Evolution of Genome Size*. John Wiley and Sons, New York.

Ceccarelli, M., Falisfocco, E. & Cionini, P.G. (1992). Variation in genome size and organization within hexaploid *Festuca arundinaceae*. *Theoret. Appl. Genet.*, 83, 273–278.

Chung, J., Lee, J.H., Arumuganathan, K., Graef, G.L. & Specht, J.E. (1998). Relationships between nuclear DNA content and seed and leaf size in Soybean. *Theoret. Appl. Genet.*, 96, 1064–1068.

Creber, H.M.C., Davies, M.S., Francis, D. & Walker, H.D. (1994). Variation in DNA C value in natural populations of *Dactylis glomerata*. *New Phytologist*, 128, 555–561.

Daly, C., Neilson, R.P. & Phillips, D.L. (1994). A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteorgy*, 33, 140–158.

Daly, C., Taylor, G. & Gibson, W. (1997). The PRISM approach to mapping precipitation and temperature. *10th Conference on Applied Climatology, American Meteorological Society, Reno, NV.,* 10–12.

Edwards, G.A. & Endrizzi, J.L. (1975). Cell size nuclear size and DNA content relationships in *Gossypium*. *Can. J. Genet. Cytol.*, 17, 181–186.

Evans, G.M., Rees, H., Snell, C.L. & Sun, S. (1972). The relation between nuclear DNA amount and the duration of the mitotic cycle. *Chromosomes Today*, 3, 24–31.

Franklin, J. (1998). Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *J. Vegetation Sci.*, 9, 733–748.

Grime, J.P. & Mowforth, M.A. (1982). Variation in genome size – an ecological interpretation. *Nature*, 299, 151–153.

Grime, J.P., Shacklock, M.L. & Band, S.R. (1985). Nuclear DNA contents, shoot phenology and species co-existence in a limestone grassland community. *New Phytologist*, 100, 435–445.

Grime, J.P., Thompson, K., Hunt, R., Hodgson, J.G., Cornelissen, J.H.C., Rorison, I.H., Hendry, G.A.F., Ashenden, T.W., Askew, A.P., Band, S.R., Booth, R.E., Bossard, C.C., Campbell, B.D., Cooper, J.E.L., Davison, A.W., Gupta, P.L., Hall, W., Hand, D.W., Hannah, M.A., Hillier, S.H., Hodkinson, D.J., Jalili, A., Liu, Z., Mackey, J.M.L., Matthews, N., Mowforth, M.A., Neal, A.M., Reader, R.J., Reiling, K., Ross-Fraser, W., Spencer, R.E., Sutton, F., Tasker, D.E. & Thorpe, P.C. (1997). Integrated screening validates primary axes of specialization in plants. *Oikos*, 79, 259–281.

Hickman, J.C. (1993). *The Jepson Manual of Higher Plants of California*. University of California Press, Berkeley, CA.

Johnston, J.S., Bennett, M.D., Rayburn, A.L., Galbraith, D.W. & Price, H.J. (1999). Reference standards for determination of DNA content of plant nuclei. *Am. J. Bot.*, 86, 609–613.

Jones, R.N. & Brown, L.M. (1976). Chromosome evolution and DNA variation in *Crepis*. *Heredity*, 36, 91–104.

Knight, C.A. & Ackerly, D.D. (2001). Correlated evolution of chloroplast heat shock protein expression in closely related plant species. *Am. J. Bot.*, 88 (3), 411–418.

Koenker, R. & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46, 33–50.

Koenker, R. & Hallock, K.F. (2001). Quantile regression: an introduction. *J. Economic Perspectives*, in press.

Koenker, R., Ng, P. & Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, 81, 673–680.

Koenker, R. & Portnoy, S. (1996). Quantile regression. University of Illinois at Urbana-Champaign, College of Commerce and Business Administration, Office of Research Working Paper 97–0100.

Levin, D.A. & Funderburg, S.W. (1979). Genome size in angiosperms: temperate versus tropical species. *Am. Naturalist*, 114, 784–795.

MacGillivray, C.W. & Grime, J.P. (1995). Genome size predicts frost resistance in British herbaceous plants: Implications for rates of vegetation response to global warming. *Func. Ecol.*, 9, 320–325.

Manly, B.F.J. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman & Hall, New York.

Mikscke, J.P. (1967). Variation in DNA content in several gymnosperms. *Can. J. Genet. Cytol.*, 9, 717–722.

Mikscke, J.P. (1971). Intraspecific variation of DNA per cell between *picea sitchensis* (Bong.) Carr. provenances. *Chromosoma*, 32, 343–352.

Mowforth, M.A. (1985). Variation in nuclear DNA amounts in flowering plants: an ecological analysis. PhD Thesis, University of Sheffield, UK.

Mowforth, M.A. & Grime, J.P. (1989). Intra-population variation in nuclear DNA amount, cell size and growth rate in *Poa annua* L. *Func. Ecol.*, 3, 289–295.

Palomino, G. & Sousa, S.M. (2000). Variation of nuclear DNA content in the biflorus species of *Lonchocarpus* (Leguminosae). *Ann. Bot.*, 85, 69–76.

Price, H.J. (1988). Nuclear DNA content variation within angiosperm species. *Evol. Trends Plants*, 2, 53–60.

Price, H.J., Chambers, K.L. & Bachmann, K. (1981). Geographic and ecological distribution of genomic DNA content variation in *Microseris douglasii* (Asteraceae). *Bot. Gazette*, 142, 415–426.

Rayburn, A.L. (1990). Genome size variation in Southwestern United States Indian maize adapted to various altitudes. *Evol. Trends Plants*, 4, 53–57.

Rayburn, A.L. & Auger, J.A. (1990). Genome size variation in *Zea mays ssp. mays* adapted to different altitudes. *Theoret. Appl. Genet.*, 79, 470–474.

Rayburn, A.L., Price, H.J., Smith, J.D. & Gold, J.R. (1985). C-band heterochromatin and DNA content in *Zea mays*. *Am. J. Bot.*, 72, 1610–1617.

Reeves, G., Francis, D., Davies, M.S., Rogers, H.J. & Hodkinson, T.R. (1998). Genome size is negatively correlated with altitude in natural populations of *Dactylis glomerata*. *Ann. Bot.*, 82 (Suppl. A), 99–105.

Scharf, F.S., Juanes, F. & Sutherland, M. (1998). Inferring ecological relationships from the edges of scatter diagrams: comparison of regression techniques. *Ecology*, 79, 448–460.

Sims, L.E. & Price, H.J. (1985). Nuclear DNA content variation in *Helianthus* (Asteraceae). *Am. J. Bot.*, 72, 1213–1219.

Sokal, R.R. & Rohlf, J.F. (1995). *Biometry*. W.H. Freeman, New York.

Stebbins, G.L. (1971). *Chromosomal Evolution in Higher Plants*. Addison-Wesley, Reading, MA.

Teoh, S.B. & Rees, H. (1976). Nuclear DNA amounts in populations of *Picea glauca* and *Pinus* species. *Heredity*, 36, 123–137.

Terrell, J.W., Cade, B.S., Carpenter, J. & Thompson, J.M. (1996). Modeling stream fish habitat limitations from wedge-shaped patterns of variation in standing stock. *Trans. Am. Fish Soc.*, 125, 104–117.

Thompson, K. (1990). Genome size, seed size and germination temperature in herbaceous angiosperms. *Evol. Trends Plants*, 4, 113–116.

Van't Hof, J. (1975). The duration of chromosomal DAN synthesis, of the mitotic cycle, and of meiosis of higher plants. In: R.C. King, eds. *Handbook of Genetics*. Plenum Press, New York.

Wakamiya, I., Newton, R.J., Johnston, S.J. & Price, J.H. (1993). Genome size and environmental factors in the genus *Pinus*. *Am. J. Bot.*, 80, 1235–1241.

Westman, W.E. (1991). Measuring realized niche spaces: climatic response of chaparral and coastal sage scrub. *Ecology*, 72, 1678–1684.

SUPPLEMENTARY MATERIAL

Table S1 Genome size, July maximum temperature, January minimum temperature, annual precipitation, and classification as native or exotic, and annual or perennial for 401 species in the California flora.