



# Global DNA cytosine methylation as an evolving trait: phylogenetic signal and correlated evolution with genome size in angiosperms

Conchita Alonso<sup>1\*</sup>, Ricardo Pérez<sup>2</sup>, Pilar Bazaga<sup>1</sup> and Carlos M. Herrera<sup>1</sup>

<sup>1</sup> Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas, Sevilla, Spain

<sup>2</sup> Instituto de Investigaciones Químicas, Centro de Investigaciones Científicas Isla de La Cartuja, Consejo Superior de Investigaciones Científicas-Universidad de Sevilla, Sevilla, Spain

## Edited by:

Zdenka Pausova, The Hospital for Sick Children, Canada

## Reviewed by:

Leon French, University of British Columbia, Canada  
Akira Kanazawa, Hokkaido University, Japan

## \*Correspondence:

Conchita Alonso, Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas, Avenida Américo Vesputio s/n, E-41092 Sevilla, Spain  
e-mail: conalo@ebd.csic.es

DNA cytosine methylation is a widespread epigenetic mechanism in eukaryotes, and plant genomes commonly are densely methylated. Genomic methylation can be associated with functional consequences such as mutational events, genomic instability or altered gene expression, but little is known on interspecific variation in global cytosine methylation in plants. In this paper, we compare global cytosine methylation estimates obtained by HPLC and use a phylogenetically-informed analytical approach to test for significance of evolutionary signatures of this trait across 54 angiosperm species in 25 families. We evaluate whether interspecific variation in global cytosine methylation is statistically related to phylogenetic distance and also whether it is evolutionarily correlated with genome size (C-value). Global cytosine methylation varied widely between species, ranging between 5.3% (*Arabidopsis*) and 39.2% (*Narcissus*). Differences between species were related to their evolutionary trajectories, as denoted by the strong phylogenetic signal underlying interspecific variation. Global cytosine methylation and genome size were evolutionarily correlated, as revealed by the significant relationship between the corresponding phylogenetically independent contrasts. On average, a ten-fold increase in genome size entailed an increase of about 10% in global cytosine methylation. Results show that global cytosine methylation is an evolving trait in angiosperms whose evolutionary trajectory is significantly linked to changes in genome size, and suggest that the evolutionary implications of epigenetic mechanisms are likely to vary between plant lineages.

**Keywords:** angiosperms, C-value, correlated evolution, DNA cytosine methylation, epigenetics, genome size, HPLC, phylogenetic signal

## INTRODUCTION

Epigenetic regulation involves DNA methylation, histone modifications and chromatin remodeling that affect many important cell functions, including regulation of gene expression and maintenance of genomic integrity (Finnegan et al., 1998b; Bender, 2004; Grant-Downton and Dickinson, 2005; Federoff, 2012). In higher plants, methylation of cytosine residues is the chief mechanism for epigenetic modification of DNA (Finnegan et al., 1998b; Jablonka and Raz, 2009). DNA methylation involves several families of plant methyltransferases, each one introducing methyl groups at specific sequences or contributing to maintain symmetric methylation after DNA replication, and plays decisive roles in plant growth and development (Finnegan et al., 1998b, 2000). Methylated cytosines occur at variable local densities throughout genic and intergenic regions (particularly in transposable elements) of nuclear plant genomes (Cokus et al., 2008; Lister et al., 2008; Zhong et al., 2013). Within species, variations in pattern (distribution across specific sites or regions) and level (proportion of total cytosines that are methylated) of cytosine methylation may induce changes in features that are important

for individual fitness, including size, flowering phenology, fecundity, inbreeding depression and response to herbivory (Sano et al., 1990; Finnegan et al., 1998a; Johannes et al., 2009; Verhoeven et al., 2010; Herrera and Bazaga, 2011; Vergeer et al., 2012). Since pattern and level of cytosine methylation often are transgenerationally heritable (Jablonka and Raz, 2009; Verhoeven et al., 2010; Herrera et al., 2013), these findings have lent considerable support to hypotheses of epigenetically-driven microevolutionary change in plant populations (Sano et al., 1990; Jablonka and Raz, 2009; Paun et al., 2010; Herrera et al., 2014).

DNA methylation has been also implicated in macroevolution through its possible effects on speciation, diversification rate and appearance of evolutionary novelty, although these suggestions have so far received limited empirical support. Methylation-mediated epigenetic mechanisms may be central to speciation through polyploidization and hybridization, two processes particularly important in plant evolution (Paun et al., 2010; Jablonka, 2013). Cytosine methylation may confer long-term selective advantage by providing an unified mechanism for setting up a variety of functions (Colot and Rossignol,

1999), and allowing evolutionary increases in gene numbers and functional complexity (Bird, 1995). Methylated cytosines mutate more frequently than unmethylated ones (Jones et al., 1992; Nachman and Crowell, 2000; Ossowski et al., 2010), hence pervasive genomic methylation may influence macroevolutionary patterns by increasing the frequency of point mutations and favoring the appearance of evolutionary novelties (McClintock, 1984; Gorelick, 2003). Alterations in pattern or level of methylation of transposable elements can influence evolutionary trajectories by inducing chromosomal rearrangements and allowing for the emergence of genetic novelty (McClintock, 1984; Feschotte and Pritham, 2007; Bonchev and Parisod, 2013). Testing these mechanistic hypotheses on the role of DNA methylation in macroevolution requires detailed comparative information on methylation patterns of specific genomic regions, a possibility currently restricted to a handful of model organisms with detailed genomic information (Zemach et al., 2010; Takuno and Gaut, 2013). Nevertheless, changes in global cytosine methylation are associated with modifications in the methylation status of specific genic and intergenic regions (Messegueur et al., 1991; Steward et al., 2002; Choi and Sano, 2007). Consequently, in absence of extensive sequence-based information on cytosine methylation patterns, the analysis of global cytosine methylation provides an initial step to evaluate the role of this epigenetic mechanism in the evolution of non-model organisms (Rozhon et al., 2008). Useful insights can be gained, for example, by investigating whether global methylation level is an evolving attribute, as implicitly assumed by proposals linking genomic methylation and evolutionary success (Bird, 1995; Colot and Rossignol, 1999; Federoff, 2012). In this paper we adopt a phylogenetically-informed approach to look for evolutionary signatures of global cytosine methylation in a sample of angiosperm species.

Interspecific variation in global cytosine methylation in plants remains unexplored, although occasional remarks indicate that it may differ substantially between species (Finnegan et al., 1998b; Bender, 2004). Published estimates of global cytosine methylation refer to few species (e.g., Wagner and Capesius, 1981; Messegueur et al., 1991), are widely scattered in the literature, and have been not examined comparatively to date. It is not known, for example, whether species differences in global cytosine methylation level reflect their distinct evolutionary histories, or whether such differences are evolutionarily correlated with concomitant changes in other attributes. Genome size is an evolving, species-specific trait in angiosperms that varies over several orders of magnitude and is correlated, among other, with life history and ecological features (Bennett and Leitch, 2012; Leitch and Leitch, 2013). A substantial part of genome size variation is caused by differences in the presence and amplification of transposable elements, particularly retrotransposons (Bennetzen et al., 1998; Grover and Wendel, 2010). Transposable elements usually are the most densely methylated regions of plant genomes (Rabinowicz et al., 2003; Federoff, 2012), and methylation-based mechanisms seem to underlie the evolution of angiosperm genome size (Bird, 1995; Federoff, 2000, 2012; Grover and Wendel, 2010), hence the evolution of global methylation levels in angiosperms might be correlated with the evolution of genome size. We analyze here an assembled data set consisting of

published and unpublished estimates of global cytosine methylation for species of angiosperms, along with related information on genome size, to address the following two specific questions: (1) Is interspecific variation in global cytosine methylation phylogenetically structured?, and (2) Are global cytosine methylation and genome size evolutionarily correlated? Results support the view that, in angiosperms, global DNA cytosine methylation is an evolving, phylogenetically structured trait whose evolutionary change is related to alterations in genome size. The nature of the relationship suggests that the evolutionary significance of methylation-driven epigenetic mechanisms is likely to differ between lineages depending on genome size.

## METHODS

Because of its accuracy and reproducibility, high-performance liquid chromatography (HPLC) is considered a “gold standard” method for estimating global DNA methylation (Fraga and Esteller, 2002; Lisanti et al., 2013). We used a reversed-phase HPLC technique to estimate global cytosine methylation in genomic DNA from full-grown current season leaves of 21 species of wild-growing plants collected at several field sites in southeastern Spain (electronic Supplementary Material, Table S1). Species were chosen to enhance, as far as possible, phylogenetic diversity (number of families represented), particularly when genome size information was available in public data bases (see below). Two different samples, each from a distinct individual, were analyzed per species. Total genomic DNA was extracted from dry leaf samples using Qiagen DNeasy Plant Mini Kit. A 100 ng aliquot was digested with 3 U of DNA Degradase Plus™ (Zymo Research, Irvine, CA), a nuclease mix that degrades DNA to its individual nucleoside components. Digestion was carried out in a 40 µL volume at 37°C for 3 h, and terminated by heat inactivation at 70°C for 20 min. Digested samples were stored at –20°C until analyzed. Global cytosine methylation was determined with a spectrofluorimetric detection technique modified after (Lopez Torres et al., 2011), and estimated for each sample as  $100 \times 5\text{mDC}/(5\text{mDC} + \text{dC})$ , where 5mDC and dC are the integrated areas under the peaks for 5-methyl-2'-deoxycytidine and 2'-deoxycytidine, respectively.

We screened the literature for estimates of global cytosine methylation for angiosperm species obtained by application of quantitative chromatographic methods based on HPLC techniques. Since different plant parts may differ in methylation level (Messegueur et al., 1991; Finnegan et al., 1998b), only estimates referred to leaf material were considered. In the case of experimental studies that included methylation level estimates for control and treated plants, only the values referred to control groups were taken into consideration. When multiple estimates from different strains or genotypes were reported, an average was computed. In total, 46 published estimates of global cytosine methylation level referred to 34 different species were gathered (electronic Supplementary Material, Table S1). This information and the results of our HPLC analyses for Spanish plants were combined into a single data set, and a mean figure computed for each species. One species (*Arabidopsis thaliana*) was present in the two groups of data, thus the final data set consisted of mean methylation levels for 54 species in 25 plant families (electronic Supplementary Material, Table S2).

Information on genome size of the species included in our sample was obtained from the Kew Royal Botanic Gardens Angiosperm C-value Database (Bennett and Leitch, 2012; last accessed 10 October 2013), which incrementally collates published information on plant C-values (Gregory, 2005) and is the standard source for contemporary work on genome size evolution in plants (Bennett and Leitch, 2005, 2012; Leitch and Leitch, 2013). C-value data, corresponding to the amount in picograms of DNA contained within a haploid nucleus, were available for 42 species in our sample (77.8% of total; electronic Supplementary Material, Table S2). When more than a single C-value was available for a species, the mean was used.

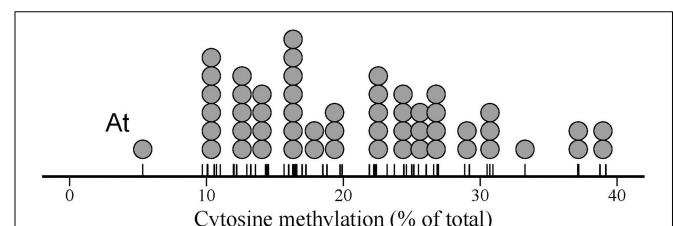
Evolutionary relationships between the 54 species included in the sample were inferred by constructing a phylogenetic tree using the phylomatic (Webb and Donoghue, 2005) tool bundled in Phylocom 4.2 software (Webb et al., 2008) in combination with the plant megatree R20120829 (available at <https://github.com/camwebb/tree-of-trees/tree/master/megatrees>; last accessed 10 October 2013). The phylogenetic tree obtained was then edited manually to eliminate polytomies (mostly in Brassicaceae and Poaceae), which were resolved by consideration of detailed family- or genus-level phylogenies available in TreeBase (<http://treebase.org>; last accessed 10 October 2013). All tree branch lengths were arbitrarily set to unity. The phylogenetic tree used in the analyses of phylogenetic signal and correlated evolution between cytosine methylation and genome size is presented in Figure S1 of the electronic Supplementary Material.

The relationship between phylogeny and interspecific variation in global cytosine methylation in our sample was examined by testing for the presence of a phylogenetic signal in the data, defined as “a tendency for related species to resemble each other more than they resemble species drawn at random from the tree” (Blomberg and Garland, 2002). Different indices have been proposed to test for phylogenetic signal in quantitative traits, which differ in performance, sensitivity to true underlying patterns of phylogenetic signal, robustness to phylogeny size, and degree of resolution of tree structure, but are robust to missing branch length (Münkemüller et al., 2012). In order to account for these differences, four indices possessing complementary features were applied concurrently to test for the presence of a phylogenetic signal in our cytosine methylation data set, namely Moran’s *I*, Abouheif’s  $C_{\text{mean}}$ , Pagel’s  $\lambda$ , and Blomberg’s *K* (Münkemüller et al., 2012). Moran’s *I* and Abouheif’s  $C_{\text{mean}}$  are autocorrelation indices that are not based on an evolutionary model and are unable to provide information on the strength of the phylogenetic signal. Blomberg’s *K* and Pagel’s  $\lambda$  assume a Brownian Motion (BM) model of trait evolution, and for both indices values close to zero denote phylogenetic independence and a value of unity indicates a trait distribution as expected under BM. Blomberg’s *K* and Pagel’s  $\lambda$  can therefore be used to assess the strength, or “effect size,” of phylogenetic structuring (see Münkemüller et al., 2012 for details on the indices used, including simulation-based comparisons of performance and limitations). Computations were performed with functions in the packages *ade4* (Jombart et al., 2010) and *phytools* (Revell, 2012) for the R environment (R Core Team et al., 2013). Statistical significance was tested in all cases by randomization with  $10^5$  repetitions.

Correlated evolution between global cytosine methylation and genome size was tested in the subset of 42 species with complete data. The phylogenetically independent contrasts (PICs) method was used, which takes into account the statistical non-independence of data due to the phylogenetic relationships between species (Felsenstein, 1985; Pagel, 1999). Genome size data were  $\log_{10}$ -transformed for analyses. Contrasts were obtained with the *pic* function in the *ape* package (Popescu et al., 2012). Departures from the BM model underlying the PICs method (continuous traits evolve randomly in any direction and amounts of change are normally distributed) may contribute to inflate Type I error (Díaz-Uriarte and Garland, 1996). Normality of contrasts, a condition expected under BM (Paradis, 2012), was tested with the Shapiro-Wilk *W* test. Significance of the relationship between the contrasts for cytosine methylation and genome size ( $\log_{10}$ -transformed) was tested by fitting a linear regression through the origin and testing its significance with a permutation procedure (*lmorigin* function in *ape* package; Paradis, 2012; Popescu et al., 2012). Conventional regression diagnostics were also performed, including checks for linearity, normality of residuals and homogeneity of variance (all adjusted for phylogeny) (Freckleton, 2009).

## RESULTS

Species means for global cytosine methylation (%mC hereafter) varied widely in our 54-species sample (Figure 1), ranging between 5.3% (*Arabidopsis thaliana*) and 39.2% (*Narcissus nevadensis*) (interquartile range = 14.3–26.5 %; mean  $\pm$  SE =  $20.8 \pm 1.1$ ,  $n = 54$ ). *Arabidopsis thaliana* stood out of the rest with a %mC value that was 1.9 standard deviations smaller than the overall sample mean (Figure 1). The broad interspecific range in %mC was not the spurious consequence of combining estimates obtained in distinct laboratories with different HPLC protocols. Mean %mC for estimates from the literature (mean  $\pm$  SE =  $21.9 \pm 1.4$  %,  $n = 46$ ) was nearly identical to the mean value for the estimates obtained for this study (mean  $\pm$  SE =  $20.4 \pm 1.5$  %,  $n = 41$ ) ( $F_{1, 85} = 0.53$ ,  $p = 0.47$ ). Within each of these two groups of data, species with two or more replicate measurements differed significantly in %mC, and interspecific variation was by far the major source of sample-wide variance ( $F_{5, 12} = 35.92$ ,  $p \ll 0.001$ , adjusted  $R^2 = 0.91$ , for literature



**FIGURE 1 |** Dot plot of global DNA cytosine methylation in the sample of 54 angiosperm species studied. Each dot represents one species, and dot width corresponds to the maximum bin width as determined from the data using a dot-density binning algorithm. The position of data points on the axis are denoted by tick marks. “At” indicates the position of *Arabidopsis thaliana*, discussed in the text.

data;  $F_{19, 20} = 53.80$ ,  $p \ll 0.001$ , adjusted  $R^2 = 0.96$ , for estimates obtained for this study). Given the contrasting sources involved in the two data groups, the close similarity in proportion of variance explained by interspecific differences further rules out methodological heterogeneity as an influential source of %mC variance in the combined sample.

All indices used to test for a phylogenetic signal in the %mC data set yielded statistically significant results (Table 1), thus strongly supporting the view that interspecific variation in global methylation level was phylogenetically structured in our species sample (Figure 2). Pagel's  $\lambda$  estimate for our data (0.877) is close to unity, the expected value for trait evolution under a pure BM model. These findings for the whole, combined sample were corroborated when literature data only were analyzed separately, and resulted much stronger when data obtained in our lab following a better hierarchical sampling of angiosperm families were analyzed separately (Table S3). Also,  $C$ -value exhibited a strong phylogenetic signal under BM evolutionary model (Pagel's  $\lambda = 0.999$ ) but moderate phylogenetic autocorrelation (Table 1, Figure 2) in our moderate sample of species.

Phylogenetic contrasts for cytosine methylation (PIC %mC) and ( $\log_{10}$ ) genome size (PIC logCvalue) did not depart significantly from normality ( $W = 0.98$  and  $0.97$ ,  $p = 0.54$  and  $0.41$ , respectively; Shapiro-Wilk tests). The regression through the origin of PIC %mC on PIC logCvalue was highly significant (adjusted  $R^2 = 0.376$ ,  $F_{1, 40} = 25.66$ ,  $p = 0.00001$ ). Regression residuals did not depart significantly from normality ( $W = 0.98$ ,  $p = 0.67$ ), and visual inspection of the plot of residuals versus predicted values did not reveal obvious departures from linearity or homocedasticity. The relationship between PIC %mC and PIC logCvalue was a direct one, revealing a correlated evolution of global cytosine methylation and ( $\log$ ) genome size in our species sample (Figure 3). The slope ( $\pm$ SE) of the fitted regression was  $9.0 \pm 1.8$ , indicating that an unity increase in  $\log_{10}$  C-value (i.e., a 10-fold increase in genome size) was evolutionarily associated with an average increase of  $\sim 10\%$  in global cytosine methylation. The bivariate plot of the raw, species-level data for %mC and logCvalue (uncorrected for phylogeny) reveals that the range of two orders of magnitude in C-value represented in our sample roughly corresponds to a range of 10–40% in %mC (Figure 4). The model plant *Arabidopsis thaliana* occupies an

outlying position in the plot, with a genome simultaneously characterized by its small size and low methylation level in comparison to the rest of species considered.

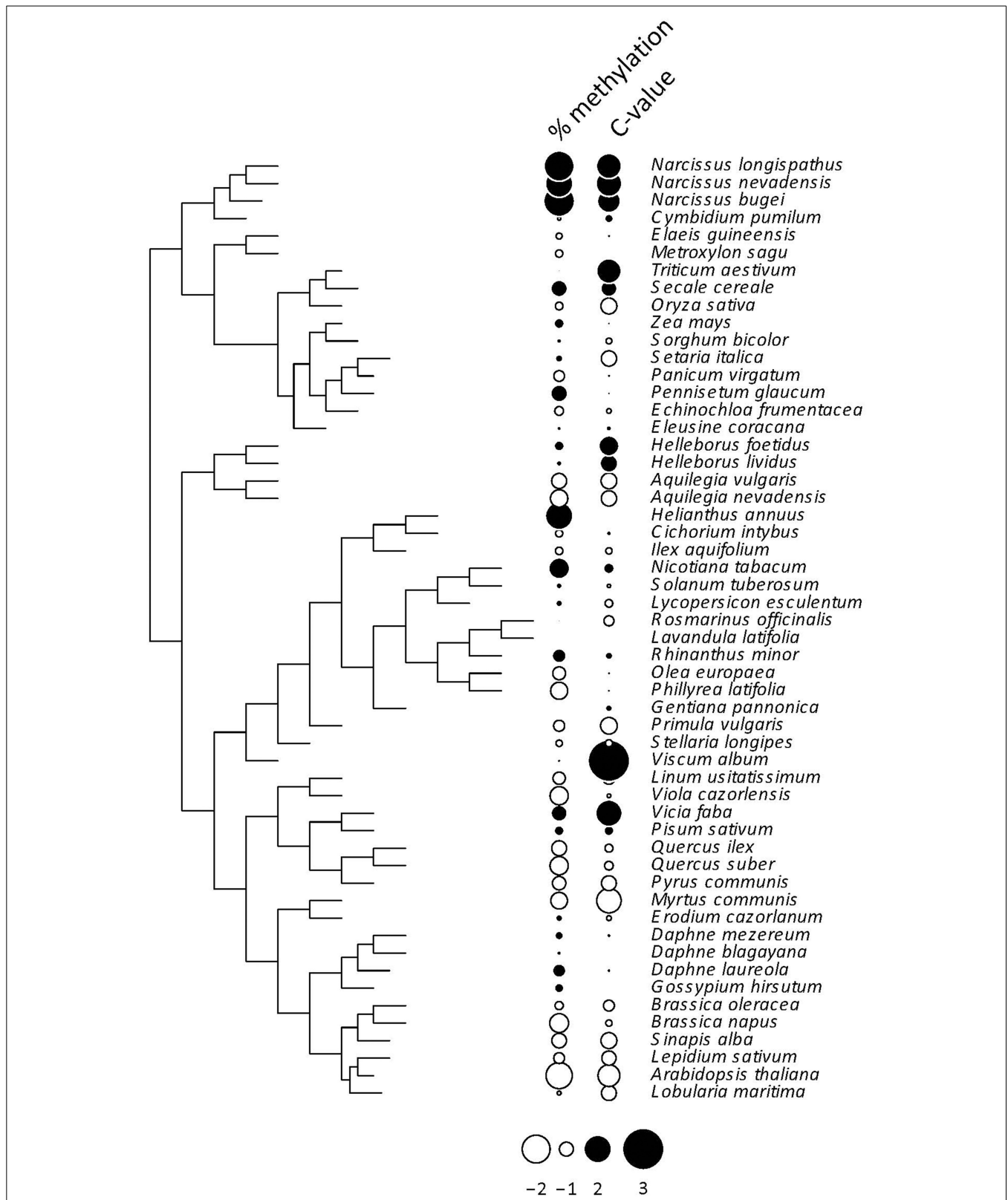
## DISCUSSION

Although global cytosine methylation measurements do not provide information on the genomic positions at which methylation occurs, it is still a valuable parameter because genome-wide methylation level can be associated with functional consequences such as mutational events, genomic instability, altered gene expression or chromosomal rearrangements (McClintock, 1984; Steward et al., 2002; Feschotte and Pritham, 2007; Rozhon et al., 2008; Bonchev and Parisod, 2013). On average, species of angiosperms considered in this study had  $\sim 20\%$  of their genomic cytosines methylated. There was, however, considerable spread around this mean value, a result that confirms earlier suggestions of interspecific variability (Wagner and Capesius, 1981; Messegueur et al., 1991) in a larger and taxonomically diverse sample. More importantly, our results demonstrate for the first time that global cytosine methylation is an evolving trait in angiosperms and differences between species are related to their evolutionary trajectories, as revealed by the strong phylogenetic signal underlying interspecific variation.

Phylogenetic signal, the tendency for evolutionarily related species to resemble each other, is ubiquitous and has been documented for a myriad morphological, physiological, life history, behavioral and ecological traits (Blomberg et al., 2003). Simulations have shown that the ability to infer evolutionary process from the measurement of phylogenetic signal alone is limited, since different evolutionary processes can produce similar phylogenetic signal, and similar evolutionary processes can eventually lead to contrasting phylogenetic signal signatures (Revell et al., 2008). These caveats, however, apply particularly to situations where observed phylogenetic signal is low (Revell et al., 2008). In our case, both Pagel's  $\lambda$  and Blomberg's  $K$  revealed a strong phylogenetic signal in %mC data, and highly significant Moran's  $I$  and Abouheif's  $C_{\text{mean}}$  confirmed phylogenetic autocorrelation in %mC data. Pagel's  $\lambda$  is a scaling parameter for the correlations between species relative to the correlation expected under Brownian evolution, and its prevailing linear relationship with strength of Brownian motion renders it a suitable index to measure strength of phylogenetic signal (Münkemüller et al., 2012). Our  $\lambda$  estimate of 0.88 for %mC data and 0.99 for C-value are very close to the parameter's practical upper limit of unity, thus indicating a strong phylogenetic signal. In contrast to  $\lambda$ , Blomberg's  $K$  does not vary linearly with strength of Brownian motion, and its upper limit depends on the number of species in the phylogeny (Blomberg et al., 2003; Münkemüller et al., 2012). Judging from Figure 2 in Münkemüller et al. (2012), the  $K$ -value obtained here for %mC data and C-values (0.39 and 0.41, respectively) fall around the inferred upper limit for our number of species sampled, which likewise supports a strong phylogenetic signal in our %mC data set. With the due caution, therefore, we suggest that the high phylogenetic signal exhibited by interspecific variation in global cytosine methylation in our species sample should be interpreted as an indication of the traits' stochastic evolution along the hierarchical tree (Figure 2). Regarding C-value,

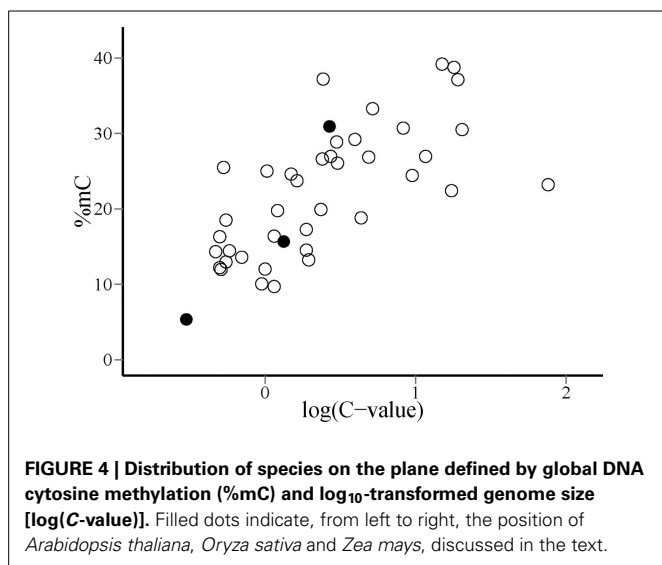
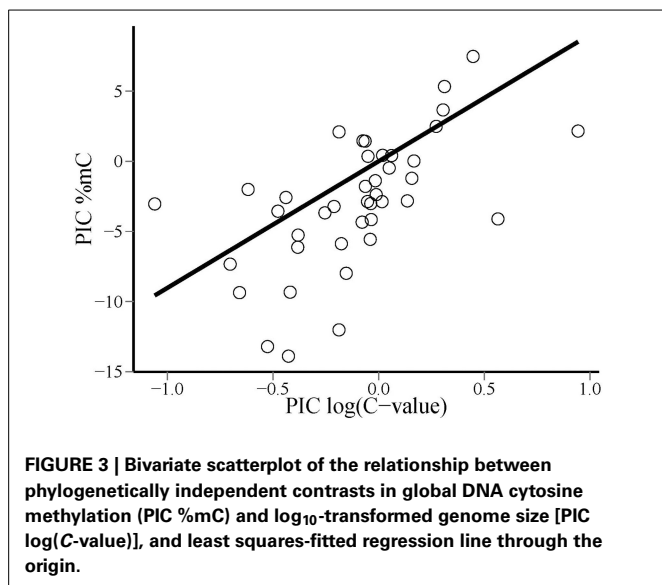
**Table 1 | Tests for the presence of a phylogenetic signal in global DNA cytosine methylation (percent of total cytosines that are methylated) and genome size (C-value) in the sample of angiosperm species studied. Statistical significance was evaluated with randomization tests.**

Phylogenetic signal index	Global DNA cytosine methylation ( $N = 54$ species)		C-value ( $N = 42$ species)	
	Statistic	$p$ -value	Statistic	$p$ -value
Moran's $I$	0.037	0.0071	-0.008	0.16
Abouheif's $C_{\text{mean}}$	0.319	0.0008	0.106	0.078
Blomberg's $K$	0.392	0.0006	0.411	0.015
Pagel's $\lambda$	0.877	0.0003	0.999	0.0002



**FIGURE 2 | Phylogenetic tree depicting the inferred evolutionary relationships between the 54 angiosperm species considered in this study.** Information on percent cytosine methylation and C-value for each species is coded as dots beside species

names. The two variables were centered and scaled for the plot. See Tables S1, S2 for raw data, and Figure S1 for the distribution over the phylogenetic tree of the 12 species with missing C-value data.



our results are congruent with a handful of studies that, using one method or another, have shown a distinct phylogenetic signature of this trait in different samples (e.g., Zonneveld, 2008; Vesely et al., 2013; Kang et al., 2014).

Phylogenetic signal only describes a pattern whereby evolutionarily related organisms resemble each other more closely than unrelated ones, without further immediate implications as to the actual mechanisms that may have caused the resemblance. From a practical viewpoint, however, our demonstration of a phylogenetic signal in global cytosine methylation has two important consequences for future comparative studies on DNA methylation in angiosperms: the genome methylation level of unstudied species may be approximately predicted from knowledge of phylogenetic position and values for close relatives, and phylogenetic relationships among species should be taken into consideration in any comparative analysis relating cytosine methylation level

to other species traits (Felsenstein, 1985; Pagel, 1999; Blomberg et al., 2003). From a conceptual perspective, the phylogenetic signal exhibited by global cytosine methylation provides grounds for a specific, testable hypothesis bearing on the macroevolutionary significance of epigenetic mechanisms. Given the increased mutation rates caused by the methylation of cytosines (Jones et al., 1992; Nachman and Crowell, 2000; Ossowski et al., 2010) and the crucial role of mutation in speciation (Nei, 2013), a direct correlation should be expected between global cytosine methylation and speciation rate in angiosperm lineages. Whether such evolution would occur by appearance and/or changes in the activity of different methylation systems, by threshold changes in gene body methylations or a combination of these and other potential mechanisms is yet unclear and deserves further studies (Zemach et al., 2010; Takuno and Gaut, 2013). The fact that cytosine methylation is most frequent in non-coding, intergenic regions such as transposons (Rabinowicz et al., 2003, 2005; Vaughn et al., 2007) does not conflict with this hypothesized evolutionary effect. Transposons are not exempt from the increased mutation rates caused by cytosine methylation (SanMiguel et al., 1998; Ossowski et al., 2010), and such alterations could likewise generate evolutionary novelties through modification of genome structure, gene sequences or gene regulatory functions (Banks and Federoff, 1989; Feschotte and Pritham, 2007; Martin et al., 2009; Bonchev and Parisod, 2013).

In contrast with the limited evolutionary information furnished by phylogenetic signal alone, correlations between traits in comparative studies are powerful tools to identify evolutionary processes underlying variation in a trait of interest (Harvey and Pagel, 1991; Pagel, 1999). A strong evolutionary correlation existed in our species sample between genome size and global cytosine methylation, as denoted by the high coefficient of determination ( $R^2$ ) of the fitted regression through the origin between PIC %mC and PIC  $\log C\text{-value}$  (Freckleton, 2009). Over the limited phylogenetic domain considered here, evolutionary changes in genome size were significantly associated with parallel modifications in global cytosine methylation as estimated by %mC. This finding is in accordance with the hypothesis postulating that methylation-based epigenetic mechanisms have played a determinant role in allowing the evolutionary increase in size and complexity of plant genomes. More specifically, it supports the proposal that “it is precisely the elaboration of epigenetic mechanisms from their prokaryotic origins as suppressors of genetic exchanges that underlies both the genome expansion and the proliferation of TEs characteristic of higher eukaryotes” (Federoff, 2012). Genome size bears only a weak relationship to gene number in plants, and transposition of repeated elements is a major cause of plant genome expansion (Bennetzen et al., 1998; SanMiguel et al., 1998). Genes do not occupy similar amounts of genomic DNA and are not packed at the same density in genomes of different sizes. Transposable elements comprise >50% of large plant genomes (e.g., maize) but <10% of small ones (e.g., *Arabidopsis*) (Bennetzen et al., 1998; Federoff, 2012). Since cytosines usually are much more densely methylated in transposons than in genes (Rabinowicz et al., 2003, 2005; Vaughn et al., 2007), then the overall %mC content of a plant genome should be directly related to its repeated sequence

content (Bender, 2004). Therefore, given the known role of cytosine methylation in the control of transposon multiplication (Finnegan et al., 1998b), the correlated evolution of genome size and %mC found here supports the notion that cytosine methylation is an adaptive feature related to the evolutionary increase in plant genome arising from the multiplication of repeated elements (Federoff, 2012).

Nonlinearity of the relationship between changes in %mC and genome size suggests, however, that the evolutionary connection between the two variables is probably more complex than the simple cause-effect, proportionality relationship suggested above. The range of *C*-values represented in our species sample (0.24–76; electronic Supplementary Material, Table S2) spans ~300-fold variation, which largely overlaps the known range of variation for the angiosperms as a whole (Bennett and Leitch, 2005). The proportion of methylated cytosines increased monotonously over the sampled range of genome sizes, but linearity of the relationship between %mC and the log<sub>10</sub> transform of *C*-value (Figure 3) reveals that methylation level increases at a slower pace than genome size. A ten-fold increase in genome size entailed an average increase of 10% in the proportion of methylated cytosines, which means that the probability of methylation of individual cytosines declined steadily as genomes grew larger. This pattern might be caused, among other, by a decline in the proportion of repeated elements in the genome or in the methylation density of repeated elements. Available information on the genomes of *Arabidopsis thaliana*, *Oryza sativa* and *Zea mays* tends to favor the second of these mechanisms. In these species, which exemplify the direct relationship between %mC and log(*C*-value) over a broad range of genome sizes (Figure 4), total DNA repeats account for 5, 31, and 66% of the genome, respectively (Le et al., 2000; Haberer et al., 2005), while the corresponding values for %mC are 5, 16, and 27%. These figures are compatible with the interpretation that methylation density of repeated elements tends to decline as they become proportionally more important in increasing genomes. Since the genetic activity of plant transposons depends closely on their methylation status, a decline in their relative methylation with increasing genome size could enhance the opportunities for genetic, phenotypic or evolutionary effects of transposable elements in plants with larger genomes (Banks and Federoff, 1989; Bonchev and Parisod, 2013). If this speculation were substantiated by future studies, then current epigenetic and epigenomic paradigms based on functional analyses of the small-genome, low-methylation, low-transposon model plant *Arabidopsis thaliana* (Schmitz and Ecker, 2012) might perhaps require some adjustments to be extrapolated to the generality of angiosperms. In any case, our results suggest the intriguing possibility that the evolutionary significance of methylation-driven, transposon-related epigenetic phenomena is likely to differ between angiosperm lineages depending on genome size.

## CONCLUSIONS

Cytosine methylation is an important epigenetic mechanism that in plants plays significant roles in the regulation of gene expression, control of genomic integrity, individual development and growth, and response to biotic and abiotic stresses (Finnegan

et al., 1998b; Bender, 2004; Grant-Downton and Dickinson, 2005; Federoff, 2012). Here we show that global cytosine methylation varied widely between plant species and such variation is phylogenetically structured in angiosperms. The correlated evolution between global cytosine methylation and genome size supports the notion that cytosine methylation is an adaptive feature allowing the evolutionary increase in size and complexity of plant genomes.

## AUTHOR CONTRIBUTIONS

Conchita Alonso and Carlos M. Herrera participated in study design, sample collection, chemical and statistical analysis, and manuscript preparation. Ricardo Pérez developed the HPLC method and ran all the samples. Pilar Bazaga prepared DNA samples and enzymatic digestions, and participated in method refinement and sample processing. Carlos M. Herrera conceived the study and obtained funding. All authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

We thank Esmeralda López for laboratory assistance; Mónica Medrano for discussion and comments on the manuscript; Xavier Picó for *Arabidopsis* material; Olatz Mitxelena and Julio Rodríguez, Waters Cromatografía SA, Barcelona, for technical advice; two anonymous reviewers for criticisms; and Consejería de Medio Ambiente, Junta de Andalucía, for plant collection permits. Financial support for this study was provided by Estación Biológica de Doñana, EBD-CSIC, Sevilla, through the Severo Ochoa Program for Centres of Excellence (SEV-2012-0262).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fgene.2015.00004/abstract>

## REFERENCES

- Banks, J., and Federoff, N. (1989). Patterns of developmental and heritable change in methylation of the suppressor-mutator transposable element. *Dev. Genet.* 10, 425–437. doi: 10.1002/dvg.1020100604
- Bender, J. (2004). DNA methylation and epigenetics. *Ann. Rev. Plant Biol.* 55, 41–68. doi: 10.1146/annurev.arplant.55.031903.141641
- Bennett, M., and Leitch, I. (2005). “Genome size evolution in plants,” in *The Evolution of the Genome*, ed T. R. Gregory (Amsterdam: Elsevier), 89–162.
- Bennett, M., and Leitch, I. (2012). *Angiosperm DNA C-values Database (Release 8.0, December 2012)*. Available online at: <http://www.kew.org/cvalues/>
- Bennetzen, J., Sanmiguel, P., Chen, M., Tikhonov, A., Francki, M., and Avramova, Z. (1998). Grass genomes. *Proc. Natl. Acad. Sci. U.S.A.* 95, 1975–1978. doi: 10.1073/pnas.95.5.1975
- Bird, A. (1995). Gene number, noise reduction and biological complexity. *Trends Genet.* 11, 94–100. doi: 10.1016/S0168-9525(00)89009-5
- Blomberg, S., and Garland, T. (2002). Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. *J. Evol. Biol.* 15, 899–910. doi: 10.1046/j.1420-9101.2002.00472.x
- Blomberg, S., Garland, T., and Ives, A. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57, 717–745. doi: 10.1554/0014-3820(2003)057[0717:TFPSIC]2.0.CO;2
- Bonchev, G., and Parisod, C. (2013). Transposable elements and microevolutionary changes in natural populations. *Mol. Ecol. Res.* 13, 765–775. doi: 10.1111/1755-0998.12133
- Choi, C., and Sano, H. (2007). Abiotic-stress induces demethylation and transcriptional activation of a gene encoding a glycerophosphodiesterase-like protein in tobacco plants. *Mol. Genet. Genomics* 277, 589–600. doi: 10.1007/s00438-007-0209-1

- Cokus, S., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C., et al. (2008). Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452, 215–219. doi: 10.1038/nature06745
- Colot, V., and Rossignol, J. (1999). Eukaryotic DNA methylation as an evolutionary device. *Bioessays* 21, 402–411. doi: 10.1002/(SICI)1521-1878(199905)21:5%3C402::AID-BIES7%3E3.3.CO;2-2
- Díaz-Uriarte, R., and Garland, T. (1996). Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. *Syst. Biol.* 45, 27–47. doi: 10.1093/sysbio/45.1.27
- Federoff, N. (2000). Transposons and genome evolution in plants. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7002–7007. doi: 10.1073/pnas.97.13.7002
- Federoff, N. (2012). Transposable elements, epigenetics, and genome evolution. *Science* 338, 758–767. doi: 10.1126/science.338.6108.758
- Felsenstein, J. (1985). Phylogenies and the comparative method. *Am. Nat.* 125, 1–15. doi: 10.1086/284325
- Feschotte, C., and Pritham, E. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* 41, 331–368. doi: 10.1146/annurev.genet.40.110405.090448
- Finnegan, E., Genger, R., Kovac, K., Peacock, W., and Dennis, E. (1998a). DNA methylation and the promotion of flowering by vernalization. *Proc. Natl. Acad. Sci. U.S.A.* 95, 5824–5829. doi: 10.1073/pnas.95.10.5824
- Finnegan, E., Genger, R., Peacock, W., and Dennis, E. (1998b). DNA methylation in plants. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 49, 223–247. doi: 10.1146/annurev.arplant.49.1.223
- Finnegan, E., Peacock, W., and Dennis, E. (2000). DNA methylation, a key regulator of plant development and other processes. *Curr. Opin. Genet. Dev.* 10, 217–223. doi: 10.1016/S0959-437X(00)00061-7
- Fraga, M., and Esteller, M. (2002). DNA methylation: a profile of methods and applications. *Biotechniques* 33, 632–649.
- Freckleton, R. (2009). The seven deadly sins of comparative analysis. *J. Evol. Biol.* 22, 1367–1375. doi: 10.1111/j.1420-9101.2009.01757.x
- Gorelick, R. (2003). Evolution of dioecy and sex chromosomes via methylation driving Muller's ratchet. *Biol. J. Linn. Soc.* 80, 353–368. doi: 10.1046/j.1095-8312.2003.00244.x
- Grant-Downton, R., and Dickinson, H. (2005). Epigenetics and its implications for plant biology. 1. The epigenetic network in plants. *Ann. Bot.* 96, 1143–1164. doi: 10.1093/aob/mci273
- Gregory, T. (2005). The C-value enigma in plants and animals: a review of parallels and an appeal for partnership. *Ann. Bot.* 95, 133–146. doi: 10.1093/aob/mci009
- Grover, C., and Wendel, J. (2010). Recent insights into mechanisms of genome size change in plants. *J. Bot.* 2010:382732. doi: 10.1155/2010/382732
- Haberer, G., Young, S., Bharti, A., Gundlach, H., Raymond, C., Fuks, G., et al. (2005). Structure and architecture of the maize genome. *Plant Physiol* 139, 1612–1624. doi: 10.1104/pp.105.068718
- Harvey, P., and Pagel, M. (1991). *The Comparative Method in Evolutionary Biology*. Oxford: Oxford Univ. Press.
- Herrera, C. M., and Bazaga, P. (2011). Untangling individual variation in natural populations: ecological, genetic and epigenetic correlates of long-term inequality in herbivory. *Mol. Ecol.* 20, 1675–1688. doi: 10.1111/j.1365-294X.2011.05026.x
- Herrera, C. M., Medrano, M., and Bazaga, P. (2013). Epigenetic differentiation persists after male gametogenesis in natural populations of the perennial herb *Helleborus foetidus* (Ranunculaceae). *PLoS ONE* 8:e70730. doi: 10.1371/journal.pone.0070730
- Herrera, C. M., Medrano, M., and Bazaga, P. (2014). Variation in DNA methylation transmissibility, genetic heterogeneity and fecundity-related traits in natural populations of the perennial herb *Helleborus foetidus*. *Mol. Ecol.* 23, 1085–1095. doi: 10.1111/mec.12679
- Jablunka, E. (2013). Epigenetic inheritance and plasticity: the responsive germline. *Prog. Biophys. Mol. Biol.* 111, 99–107. doi: 10.1016/j.pbiomolbio.2012.08.014
- Jablunka, E., and Raz, G. (2009). Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q. Rev. Biol.* 84, 131–176. doi: 10.1086/598822
- Johannes, F., Porcher, E., Teixeira, F., Saliba-Colombani, V., Simon, M., Agier, N., et al. (2009). Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genetics* 5:e1000530. doi: 10.1371/journal.pgen.1000530
- Jombart, T., Balloux, F., and Dray, S. (2010). adephylo: new tools for investigating the phylogenetic signal in biological traits. *Bioinformatics* 26, 1907–1909. doi: 10.1093/bioinformatics/btq292
- Jones, P., Rideout, W., Shen, J., Spruck, C., and Tsai, Y. (1992). Methylation, mutation and cancer. *Bioessays* 14, 33–36. doi: 10.1002/bies.950140107
- Kang, M., Tao, J., Wang, J., Ren, C., Qi, Q., Xiang, Q.-Y., et al. (2014). Adaptive and nonadaptive genome size evolution in Karst endemic flora of China. *New Phytol.* 202, 1371–1381. doi: 10.1111/nph.12726
- Le, Q., Wright, S., Yu, Z., and Bureau, T. (2000). Transposon diversity in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7376–7381. doi: 10.1073/pnas.97.13.7376
- Leitch, I., and Leitch, A. (2013). “Genome size diversity and evolution in land plants” in *Plant Genome Diversity*, Vol. 2. *Physical Structure, Behaviour and Evolution of Plant Genomes*, eds J. Greilhuber, J. Dolezel and J. F. Wendel. (New York, NY: Springer), 307–322.
- Lisanti, S., Omar, W., Tomaszewski, B., De Prins, S., Jacobs, G., Koppen, G., et al. (2013). Comparison of methods for quantification of global DNA methylation in human cells and tissues. *PLoS ONE* 8:e79044. doi: 10.1371/journal.pone.0079044
- Lister, R., O'malley, R., Tonti-Filippini, J., Gregory, B., Berry, C., Millar, A., et al. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133, 523–536. doi: 10.1016/j.cell.2008.03.029
- Lopez Torres, A., Yanez Barrientos, E., Wrobel, K., and Wrobel, K. (2011). Selective derivatization of cytosine and methylcytosine moieties with 2-bromoacetophenone for submicrogram DNA methylation analysis by reversed phase HPLC with spectrofluorimetric detection. *Anal. Chem.* 83, 7999–8005. doi: 10.1021/ac2020799
- Martin, A., Troadec, C., Boualem, A., Rajab, M., Fernandez, R., Morin, H., et al. (2009). A transposon-induced epigenetic change leads to sex determination in melon. *Nature* 461, 1135–1138. doi: 10.1038/nature08498
- McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* 226, 792–801. doi: 10.1126/science.15739260
- Messeguer, R., Ganal, M., Steffens, J., and Tanksley, S. (1991). Characterization of the level, target sites and inheritance of cytosine methylation in tomato nuclear DNA. *Plant Mol. Biol.* 16, 753–770. doi: 10.1007/BF00015069
- Münckemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schifffers, K., et al. (2012). How to measure and test phylogenetic signal. *Methods Ecol. Evol.* 3, 743–756. doi: 10.1111/j.2041-210X.2012.00196.x
- Nachman, M., and Crowell, S. (2000). Estimate of the mutation rate per nucleotide in humans. *Genetics* 156, 297–304.
- Nei, M. (2013). *Mutation-Driven Evolution*. Oxford: Oxford University Press.
- Ossowski, S., Schneeberger, K., Lucas-Lledo, J., Warthmann, N., Clark, R., Shaw, R., et al. (2010). The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327, 92–94. doi: 10.1126/science.1180677
- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature* 401, 877–884. doi: 10.1038/44766
- Paradis, E. (2012). *Analysis of Phylogenetics and Evolution with R, 2nd Edn*. New York, NY: Springer.
- Paun, O., Bateman, R., Fay, M., Hedren, M., Civeyrel, L., and Chase, M. (2010). Stable epigenetic effects impact adaptation in allopolyploid orchids (*Dactylorhiza*: Orchidaceae). *Mol. Biol. Evol.* 27, 2465–2473. doi: 10.1093/molbev/msq150
- Popescu, A., Huber, K., and Paradis, E. (2012). ape 3.0: new tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics* 28, 1536–1537. doi: 10.1093/bioinformatics/bts184
- Rabinowicz, P., Citek, R., Budiman, M., Nunberg, A., Bedell, J., Lakey, N., et al. (2005). Differential methylation of genes and repeats in land plants. *Genome Res.* 15, 1431–1440. doi: 10.1101/gr.4100405
- Rabinowicz, P., Palmer, L., May, B., Hemann, M., Lowe, S., McCombie, R., et al. (2003). Genes and transposons are differentially methylated in plants, but not in mammals. *Genome Res.* 13, 2658–2664. doi: 10.1101/gr.1784803
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Revell, L. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223. doi: 10.1111/j.2041-210X.2011.00169.x
- Revell, L., Harmon, L., and Collar, D. (2008). Phylogenetic signal, evolutionary process, and rate. *Syst. Biol.* 57, 591–601. doi: 10.1080/10635150802302427
- Rozhon, W., Baubec, T., Mayerhofer, J., Scheid, O., and Jonak, C. (2008). Rapid quantification of global DNA methylation by isocratic cation exchange high-performance liquid chromatography. *Anal. Biochem.* 375, 354–360. doi: 10.1016/j.ab.2008.01.001



- SanMiguel, P., Gaut, B., Tikhonov, A., Nakajima, Y., and Bennetzen, J. (1998). The paleontology of intergene retrotransposons of maize. *Nat. Genet.* 20, 43–45. doi: 10.1038/1695
- Sano, H., Kamada, I., Youssefian, S., Katsumi, M., and Wabiko, H. (1990). A single treatment of rice seedlings with 5-azacytidine induces heritable dwarfism and undermethylation of genomic DNA. *Mol. Gen. Genet.* 220, 441–447. doi: 10.1007/BF00391751
- Schmitz, R., and Ecker, J. (2012). Epigenetic and epigenomic variation in *Arabidopsis thaliana*. *Trends Plant Sci.* 17, 149–154. doi: 10.1016/j.tplants.2012.01.001
- Steward, N., Ito, M., Yamaguchi, Y., Koizumi, N., and Sano, H. (2002). Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. *J. Biol. Chem.* 277, 37741–37746. doi: 10.1074/jbc.M204050200
- Takuno, S., and Gaut, B. (2013). Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1797–1802. doi: 10.1073/pnas.1215380110
- Vaughn, M., Tanurdzic, M., Lippman, Z., Jiang, H., Carrasquillo, R., Rabinowicz, P., et al. (2007). Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biol.* 5:e174. doi: 10.1371/journal.pbio.0050174
- Vergeer, P., Wagemaker, N., and Ouborg, N. (2012). Evidence for an epigenetic role in inbreeding depression. *Biol. Lett.* 8, 798–801. doi: 10.1098/rsbl.2012.0494
- Verhoeven, K., Jansen, J., Van Dijk, P., and Biere, A. (2010). Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytol.* 185, 1108–1118. doi: 10.1111/j.1469-8137.2009.03121.x
- Vesely, P., Bures, P., and Smarda, P. (2013). Nutrient reserves may allow for genome size increase: evidence from comparison of geophytes and their sister non-geophytic relatives. *Ann. Bot.* 112, 1193–1200. doi: 10.1093/aob/mct185
- Wagner, I., and Capius, I. (1981). Determination of 5-methylcytosine from plant DNA by high-performance liquid chromatography. *Biochim. Biophys. Acta* 654, 52–56. doi: 10.1016/0005-2787(81)90135-0
- Webb, C., Ackerly, D., and Kembel, S. (2008). Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics* 24, 2098–2100. doi: 10.1093/bioinformatics/btn358
- Webb, C., and Donoghue, M. (2005). Phylomatic: tree assembly for applied phylogenetics. *Mol. Ecol. Notes* 5, 181–183. doi: 10.1111/j.1471-8286.2004.00829.x
- Zemach, A., Mcdaniel, I., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328, 916–919. doi: 10.1126/science.1186366
- Zhong, S., Fei, Z., Chen, Y.-R., Zheng, Y., Huang, M., Vrebalov, J., et al. (2013). Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nat. Biotechnol.* 31, 154–159. doi: 10.1038/nbt.2462
- Zonneveld, B. J. M. (2008). The systematic value of nuclear DNA content for all species of *Narcissus* L. (*Amaryllidaceae*). *Plant Syst. Evol.* 275, 109–132. doi: 10.1007/s00606-008-0015-1

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 October 2014; accepted: 07 January 2015; published online: 29 January 2015.

Citation: Alonso C, Pérez R, Bazaga P and Herrera CM (2015) Global DNA cytosine methylation as an evolving trait: phylogenetic signal and correlated evolution with genome size in angiosperms. *Front. Genet.* 6:4. doi: 10.3389/fgenet.2015.00004

This article was submitted to *Epigenomics and Epigenetics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2015 Alonso, Pérez, Bazaga and Herrera. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.