Article (refereed) - postprint

Pandit, Maharaj K.; White, Steven M.; Pocock, Michael J.O. 2014. **The contrasting effects of genome size, chromosome number and ploidy level on plant invasiveness: a global analysis**. *New Phytologist*, 203 (2). 697-703. 10.1111/nph.12799

THE CONTRASTING EFFECTS OF GENOME SIZE, CHROMOSOME NUMBER AND PLOIDY LEVEL ON PLANT INVASIVENESS: A GLOBAL ANALYSIS

Maharaj K.Pandit[1], Steven M. White[2,3]and Michael J.O. Pocock[2]*

[1]Department of Environmental Studies & Centre for Inter-disciplinary Studies of Mountain & Hill Environment, University of Delhi, Delhi - 110007. India.

[2]Centre for Ecology & Hydrology, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK.

[3]Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, Oxfordshire, OX1 3LB, UK.

Running title: invasiveness, genome size and ploidy level

Type of article: Full research article

Word counts Abstract: 194

Main text (incl. refs): 5398

Introduction: 856

Materials and Methods: 1020

Results: 811

Discussion: 1077

Acknowledgements: 33

Number of figures: 2 (Fig. 1 currently in colour)

Number of tables: 1

Supporting information: 4 tables and 3 figures

*Author for correspondence: Michael J.O. Pocock, Centre for Ecology & Hydrology, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK. Tel: +44 (0)1491 692566. Fax: +44 (0)1491 692424. Email:michael.pocock@ceh.ac.uk

Summary

- Understanding how species' traits relate to their status (e.g. invasiveness or rarity) is important because it can help to efficiently focus conservation and management effort and infer mechanisms affecting plant status. This is particularly important for invasiveness in which pro-active action is needed to restrict the establishment of potentially invasive plants.
- We tested the ability of genome size (DNA 1C-values) to explain invasiveness and compared it to cytogenetic traits (chromosome number and ploidy level).We considered 890 species from 62 genera, from across the angiosperm phylogeny and distributed from tropical to boreal latitudes.
- We show that invasiveness was negatively related to genome size and positively related to chromosome number (and ploidy level) yet there was a positive relationship between genome size and chromosome number, i.e. our result was not due to co-linearity between the traits. Including both traits in explanatory models greatly increased the explanatory power of each.
- This demonstrates the potential unifying role that genome size, chromosome number and ploidy have as species' traits, despite the diverse impacts they have on plant physiology. It provides support for the continued cataloguing of cytogenetic traits and genome size of the world's flora.

Key words: DNA 1C-value, holoploid genome size, invasive, genomic traits, phylogenetic signal, angiosperm

INTRODUCTION

Analyses of how traits of different species relate to aspects of their status have been long considered as a tool in conservation biology (Fisher & Owens, 2004). From these relationships, it is possible to infer the mechanisms that promote or permit species' status, e.g. their rarity, invasiveness or population trends. However, while such approaches have been widely used they have had mixed success, with sometimes inconsistent results across taxonomic groups or geographic regions (Williamson & Fitter, 1996; Kunin & Gaston, 1997; Pyšek & Richardson, 2007).

Invasiveness is a trait that is especially valuable to consider with cross-species analyses because there is great value in identifying species likely to be invasive, given the huge difference in the cost of management of invasives at different stages in their establishment (Pyšek & Richardson, 2010). Of course, invasiveness is, to an extent, context-specific (van Kleunen *et al.*, 2010a). However, if invasive species could be predicted from their traits then it would support governments' efforts to fulfil their obligation to "as far as possible and as appropriate, prevent the introduction of, control or eradicate those alien species which threaten ecosystems, habitats or species" (Article 8h in the Convention on Biological Diversity (CBD)). Several biological traits have been shown to be important in explaining plant invasiveness, e.g. short generation time, high growth rate and high fitness (Pyšek & Richardson, 2007; van Kleunen *et al.*, 2010b,a; Ordonez *et al.*, 2010; Schmidt & Drake, 2011). Also species' traits such as chromosome number and ploidy level have shown potential in explaining invasiveness (Soltis & Soltis, 2000; Pandit, 2006; Pandit *et al.*, 2006, 2011). In addition to these traits, genome size has been used successfully to explain extinction risk (Vinogradov, 2003), and although it has a variable effect on invasiveness in individual taxa (Gallagher *et al.*, 2011; Varela-Álvarez *et al.*, 2012) there has been no attempt at assessing this at a large scale across the plant phylogeny.

Genome size is an invariant characteristic of an individual and usually invariant within a species; the amount of nuclear DNA follows a set of simple multiples of its basic quantity, designated as 'C-values' (1C, 2C, 4C, 8C...). 1C is the amount of DNA in the unreplicated gametic nucleus of an organism, i.e. the holoploid genome size (Greilhuber *et al.*, 2005) and the C-values have subsequently been used as a reference value for genome size studies. Nuclear DNA content varies approximately 2400-fold in angiosperms due to changes in the amount of non-coding DNA sequences and genome duplication (Bennett & Leitch, 2011). Despite what was once thought, it has no relationship with an organism's phenotypic complexity (Gregory, 2001), but it does influence a wide array of characteristics, e.g. rate of cell division, sensitivity to radiation and ecological behaviour in plant communities (reviewed in Bennett, 1987; Bennett & Smith, 1991). Genome size has been described as a trait that "uniquely lies at the intersection of phenotype and genotype" (Oliver *et al.*, 2007) and, for this reason, it has also been described as an "important biodiversity character, whose study provides a strong unifying element in biology with practical and predictive uses" (Bennett & Leitch, 2005). In plants, comparative studies have suggested that large genome size is maladaptive through its constraints on plant physiology (Vinogradov, 2003; Knight *et al.*, 2005). However, some have also suggested that large genome sizes may be beneficial, e.g. in some fish high DNA C-values (due to the accumulation of non-coding DNA) are associated with lower basal metabolic rates, which appears to allow them to adapt to environmental niches with lower energy supply (Szarski, 1983). It is also possible that variation in genome size has little adaptive value: the neutral theory of selection (Oliver *et al.*, 2007).

Genome size influences a wide range of plant physiological and evolutionary traits (Bennett & Leitch, 2005) which have individually been shown to relate to invasiveness (van Kleunen *et al.*, 2010b), so we expected that invasive plants would have relatively small genome size. This fits with the conjecture that large genome size is maladaptive (Orgel & Crick, 1980; Rejmánek, 1996).

Based on predictions of the effects of genome duplication and polyploidisation, we expected that genome size would be positively correlated with the cytogenetic traits (ploidy and chromosome number). However, we also expected a positive effect of ploidy and chromosome number on invasiveness (Pandit, 2006; Pandit *et al.*, 2006, 2011; Schmidt & Drake, 2011) because chromosome number is positively related to rates of adaptation (te Beest *et al.*, 2012) and polyploidy leads to an evolutionary advantage due to effects of heterosis and gene redundancy (Comai, 2005). The fact that these pairs of expectations are contradictory with each other was identified by Rejmánek (1996), who also identified that "research on this subject seems to be very scanty".

In the current study we tested for relationship of genome size with invasiveness in angiosperms, using a global dataset of species from across the angiosperm phylogeny. We compared these results with the relationship of cytogenetic traits (chromosome number and ploidy) with invasiveness. Throughout we considered phylogeny and the latitude of each species, given the evidence of both on genome size (Bennett *et al.*, 1998; Knight *et al.*, 2005).

MATERIALS AND METHODS

Data on chromosomal data and invasiveness

Holoploid genome size (DNA 1C-values of species in pg) and chromosome numbers were collated from the Kew Royal Botanic Gardens Plant C-values database, release 5.0 (http://data.kew.org/cvalues/; (Bennett & Leitch, 2010)). We undertook analyses on a balanced subset of the species for which there was information on genome size, ploidy level and chromosome number (described in the 'Data analysis' section below). We defined invasive plants as those that were included in the Global Invasive Species Database (GISD; http://www.issg.org/database) and Pacific Island Ecosystems at Risk (PIER; http://www.hear.org/pier/scientificnames/scinamea.htm) list. These two databases provide a global

perspective on invasiveness in plants. Our dataset therefore had similar scope and global geographic coverage to our previous study (Pandit *et al.*, 2011).

Latitudinal data

It has previously been suggested that genome size and cytogenetic traits vary according to latitude, with a peak at temperate latitudes (Bennett, 1987; Knight *et al.*, 2005). We therefore extracted information on the distribution of each species from the Global Biodiversity Information Facility (accessed through GBIF Data Portal, data.gbif.org, 2013-02-04) by calculating the average latitude of the centres of one-degree latitude/longitude grid cells in which the species had been recorded. The extraction of these data from GBIF was automated with the Rgbif package (Chamberlain *et al.*, 2012) in R 2.15.2 (R Core Team, 2012), with additional code written by us to gather data on all the synonyms of each taxon under consideration (as listed by GBIF). We considered the distribution of occupied cells rather than the distribution of individual records because it was more robust to spatial variation in recorder intensity and considered the absolute value of latitude because it provides a better assessment of the latitude for species introduced from the southern to northern hemisphere or vice versa. A small number of records may have been wrongly geo-located, but our observation of the location data suggests this is negligible in influencing the average absolute value of latitude.

Phylogenetic data

We constructed the phylogenetic tree according to a fully resolved family-level phylogeny (R20120829.new) in Phylomatic v3 (Webb & Donoghue, 2005), based on the Angiosperm Phylogeny (APG III, 2009). We calibrated the branch lengths in the tree using the BLADJ algorithm in Phylocom 4.2 (Webb *et al.*, 2008). It assigns dates to nodes contained in a dated tree (Wikström *et al.*, 2001) and then divides the remaining, unassigned, nodes evenly across time. Although simple, this is a widely-used routine that improves on alternative methods for calibrating

phylogenetic trees (Webb, 2000) and provides similar results in phylogenetically-informed analyses to other methods (e.g. Davies *et al.*, 2013). The minimum branch lengths from this analysis were 6.25 my, but because we wanted to include all aneuploids (chromosome number variants within a species; 63 instances across 52 species) in the analysis, we set their branch lengths to an arbitrary small value of 0.1 my.

Data analysis

In our analysis we tested the relationships of invasiveness with genome size and chromosome number, with and without latitude as a covariate. We found that there were computational limitations in adopting a fully phylogenetically-informed approach with the whole dataset; specifically the highly unbalanced nature of the full dataset (i.e. 90% of genera in the full dataset did not have invasive species present) regularly led to lack of model convergence, while runtime was estimated to be at least several weeks for each model (it scaled exponentially with sample size). Therefore we undertook the analysis with the 62 genera for which there were both invasive and non-invasive species. We thus excluded 854 and 35 genera for which there were, respectively, only non-invasive and invasive species, although the majority of these genera (61%) comprised only one species. We excluded a further 50 species for which distribution data was not present in GBIF, but excluding these species did not influence the final number of genera. Overall, we reduced the overall sample size from 4504 to 890 species (see Results), but we retained as many highly informative comparisons as possible (i.e. between congeners; Pandit *et al.*, 2011), while creating a smaller, more balanced dataset suitable for analysis. This then was akin to a 'sister pairs' analysis. Importantly, because species within a genus have a tendency to regionally co-occur, this analysis also helps to account for regional variation in the intensity of records in GBIF (Yesson *et al.*, 2007) and the unbalanced geographical representation of the Kew Plant C-values database (Leong-Škorničková *et al.*, 2007).

167 Given that species' traits are often not randomly distributed across phylogenetic trees, we
168 undertook analyses with a phylogenetically-informed approach, thus incorporating an appropriate
169 degree of phylogenetic signal (Revell, 2010). In our analyses when the response trait was
170 continuous, we used phylogenetic generalised least squares (PGLS) analyses using the function
171 'pgls' in 'caper' (Orme *et al.*, 2011). When the response variable was binary (e.g. invasive or not),
172 we used phylogenetic logistic regression (PLR) (Ives & Garland, 2010), which is a logistic
173 regression with the appropriate degree of phylogenetic signal, run in MATLAB (Release 2013a,
174 The MathWorks, Inc., Massachusetts) with code available from T. Garland.

175 For all analyses, we complemented the fully phylogenetically-informed approaches with a
176 generalised linear mixed model (GLMM) in which genus was treated as a random intercept, thus
177 retaining within-genus comparisons. Although reporting both phylogenetically-informed and
178 cross-species analyses is not recommended (Freckleton, 2009), the value of using GLMMs is that
179 they allowed us to assess model fit (both absolute model fit with r^2 and relative model fit with
180 Akaike's information criterion: AIC); these values are not currently possible to obtain for PLRs
181 (Ives & Garland, 2010). Model fit was apportioned as the proportion of variance explained by the
182 fixed effects ($r^2_{\mathrm{GLMM}(m)}$) and the proportion of variance explained by the total model ($r^2_{\mathrm{GLMM}(c)}$)
183 (Nakagawa & Schielzeth, 2013). These models were run with the function 'lmer' and the
184 significance of the variables were estimated with 'mcmcamp' in 'lme4' (Bates *et al.*, 2012) in R
185 2.15.2.

186 We also tested for a positive relationship between genome size and cytogenetic traits
187 (chromosome number and ploidy) by using PGLS models with genome size as the dependent
188 variable and by considering the additive and interaction effects of latitude on the relationship.

RESULTS

Our final dataset comprised the species for which we had chromosome numbers, genome size and distribution data, from all genera for which there were both invasive and non-invasive species: i.e. 890 species from 62 genera in 27 families belonging to 21 orders. The species in the dataset were from across the angiosperm phylogeny (Fig. S1) and were well distributed across latitudes, from tropical to northern temperate regions (Fig. S2).

We found that invasiveness was negatively related to holoploid genome size but positively related to chromosome number (Table 1; Figs 1 & 2). We found best support for models that included genome size and chromosome number together. In these models the qualitative results were the same as for the traits individually but the magnitude and significance of the effects was increased (Table 1; Figs 1 & 2). The models explaining invasiveness showed little phylogenetic signal (in the PLRs the measure of phylogenetic signal was low: $a < -2.7$; Ives & Garland, 2010) which is what we expected because 'invasiveness' is a complex trait that is not directly inherited. These findings confirmed our expectations, and the simplest way of explaining them is that the two independent traits are negatively associated. However, the findings were particularly striking because genome size and chromosome number are actually positively related, as we predicted (Figs 2 & S3; Table S1). This positive relationship showed strong phylogenetic signal (in the PGLS models the measure of phylogenetic signal was high: $\lambda > 0.92$; Revell, 2010) which confirmed our expectations because both genome size and chromosome number are directly inherited.

We used three lines of evidence supporting the conclusion that genome size and chromosome number are best included together in models to explain invasiveness: model fit (r^2), relative model fit (AIC) and standardised effect sizes (the latter two as recommended by Freckleton (2009)). It is not currently possible to obtain r^2 or AIC for PLRs (Ives & Garland 2010) so we relied on the

213 results of the GLMMs. We were confident in doing this because the measure of phylogenetic
214 signal in the PLRs was low (a< -2.7) and model parameters were similar between the two (Table
215 1). The fit of the fixed effects to the data ($r^2_{GLMM(m)}$) increased considerably when the two traits
216 were included together (i.e. r^2 rose from <4% with each univariate model to 9% with both traits;
217 Table 1). The best fitting candidate model (i.e. lowest AIC) was that which included both traits,
218 with some support for the model with an interaction between the two and decreasing support for
219 the models with chromosome number alone and genome size alone (Table 1). The standardised
220 model parameters revealed that standardised effect sizes of genome size and chromosome number
221 were similar in magnitude, albeit in opposite directions, but when included together the magnitude
222 of each almost doubled (Fig. 2). In other words, genome size not only explains variation in
223 invasiveness but, importantly, it explains residual variation of the relationship of chromosome
224 number with invasiveness.

225 We present results for chromosome number because this is a directly observable trait but all our
226 reported results were very similar with ploidy level (Tables S1 & S2). Latitude was not an
227 important explanatory variable for invasiveness, chromosome number or ploidy level (Table S3).
228 Genome size was significantly higher at higher latitudes but there was no evidence of a unimodal
229 (quadratic) relationship. Latitude was not an important covariate in models explaining
230 invasiveness (Table S2). There was little phylogenetic signal in the results (the value of
231 phylogenetic signal, a, in the PLR models was always < -2.7 (Tables 1 and S2; Ives & Garland,
232 2010). Also, although we used information on invasive plants from two sources (the GISD and the
233 PIER database), all our results were qualitatively similar whether considering GISD alone, PIER
234 alone or both (Table S4).

235 The simplest explanation for our findings about the relationship between genome size or
236 chromosome number and invasiveness was that the two are negatively associated, but the data

confirmed out expectations that genome size is significantly positively related to chromosome number. The simplest PGLS model was: $\log_2$ (DNA C-value) = -1.327 + (0.460 × $\log_2$(chromosome number)), with both intercept and slope being significantly different from zero (P=0.047 and P<0.001, respectively). Therefore, a doubling of chromosome number results in a 1.38-fold increase in genome size (because $2^{0.460}$ = 1.38). However there was support for a more complex PGLS model in which genome size was a function of the interaction between chromosome number and latitude squared. The relationship of genome size with chromosome number was steepest at high latitudes (a doubling of chromosome number resulted in a 1.8-fold increase in genome size when latitude was 55°, but a 1.3-fold increase at 30°; Fig. S3). In all PGLS models the effect of phylogeny was substantial (λ>0.925, indicating strong phylogenetic autocorrelation). There was a similarly strong relationship of genome size with ploidy level (Table S1).

DISCUSSION

The results presented in this study show that there is strong evidence that invasiveness is associated with both smaller genome sizes and larger chromosome numbers (and ploidy levels). The results also show that there is synergy in explaining invasiveness with both traits together rather than considering each separately. The results for the individual traits are despite the conflicting positive relationship of genome size with chromosome number (and ploidy) and so all three sets of relationships (Fig. 2) confirm the conjecture of Rejmánek (1996) using a global dataset of species from across the angiosperm phylogeny.

Our results raise two important questions. The first question is: how is it possible for all three relationships to be significant when they appear to conflict? Co-linearity between genome size and chromosome number would have been the simplest explanation, but these traits are positively

related (Fig. 2), so co-linearity is not the answer (Rejmánek, 1996). The effect of genome size and chromosome number is much stronger when considering both traits together in an analysis (i.e. standardised betas are increased; Fig. 2), which shows the important of genome size, when considering the effect of chromosome number, and vice versa. Therefore, one parsimonious interpretation is that invasiveness is related to changes in chromosome number/ploidy (and its consequent effect on genome size) and to changes in genome size for a given chromosome number/ploidy. Genome downsizing after whole genome duplication (Ibarra-Laclette *et al.*, 2013) also helps explain these effects and there may be interactions between the effects of genome size and ploidy on plant physiology, e.g. increases in genome size being more important as ploidy level increases (Bennett & Smith, 1972).

The second important question raised by the results is: what are the causal mechanisms explaining the relationship of invasiveness with genome size and chromosome number/ploidy? Genome size, chromosome number and ploidy each have effects on diverse aspects of plant physiology, and there are many mechanisms by which they may influence plant status, such as invasiveness. Considering genome size, it appears to affect adaptability of plant species, with larger genome sizes failing to adapt to variable habitats, while plants with smaller genomes, thrive successfully and become invasive (Bennett, 1987; Bennett *et al.*, 1998). This is possibly because smaller genomes are associated with smaller cell size (Cavalier-Smith, 1982) and faster rates of mitotic and meiotic divisions (Gregory, 2001; Knight & Beaulieu, 2008; Francis *et al.*, 2008), faster germination (Minelli *et al.*, 1996) and hence reduced generation times (Bennett, 1972; Grime *et al.*, 1985; Mowforth & Grime, 1989). It is likely that this is an adaptation to time-limited envionrments, so pre-adapting the plant to invasiveness (Rejmánek, 1996). Smaller genome size is also associated with smaller seed mass (Bennett, 1987; Knight & Ackerly, 2002) and lower plant height (Minelli *et al.*, 1996), which due to complex trade-offs in plant traits could lead to

increased or decreased spread of spread and competetiveness (Thomson *et al.*, 2011; Caplat *et al.*, 2012). Even stronger evidence for these mechanisms comes from within-species studies, e.g. that genome downsizing leads to increased colonization potential (Lavergne *et al.*, 2010). Polyploidy, and hence higher chromosome numbers, also contribute to increase invasiveness through the beneficial effects of heterosis, increased speed of cell division, gene redundancy and increased phenotypic variation (Bennett & Smith, 1972; Comai, 2005; te Beest *et al.*, 2012) which can 'pre-adapt' taxa to be invasive or to evolve invasiveness (te Beest *et al.*, 2012). Empirical studies on individual invasive plant species such as *Centaurea stoebe* (=*C. maculosa*) (Treier *et al.*, 2009; Hahn *et al.*, 2012) and *Claytonia perfoliata* (McIntyre, 2012) help elucidate these mechanisms and they have been discussed in previous cross-species studies on the effect of chromosome number and ploidy on plant status (Pandit, 2006; Pandit *et al.*, 2011).

We found no effect of latitude on the relationship of chromosomal traits with invasiveness (Table S2), but genome size increases with latitude, when taking chromosome number into account, and it increases more rapidly with chromosome number at higher latitudes (Table S1; Fig. S3). This relationship appeared linear rather than unimodal (Bennett *et al.*, 1998; Knight *et al.*, 2005) probably because we had few high latitude species in the dataset (the absolute latitude of the range of most species was < 60°) and the omission of arctic species may explain the lack of an observed relationship of latitude with ploidy.

Plant traits such as genome size, ploidy and chromosome number show potential to be unifying characters explaining plant status, but we believe that there is important future work to further elucidate the mechanisms linking these traits to invasiveness and to discover how these relate to the different stages in the route to becoming invasive (Kubešová *et al.*, 2010). Within this context, the intention to continue cataloguing the genome size of the world's flora (Galbraith *et al.*, 2011; Bennett & Leitch, 2011) is to be welcomed. We note, however, that increasing representation of

species within genera, where arguably it is most useful in conservation practice, is not a specific target of the Plant Genome Size workshops (Bennett & Leitch, 2011). Despite holoploid genome size being "less cumbersome" to measure than chromosome number (Galbraith *et al*., 2011), our results show that both traits are important and data on both traits should be collected for maximum benefits to conservation practice.

Finally, the bigger evolutionary question that needs to be answered is the role and existence of 'selfish' DNA (Orgel & Crick, 1980). Whether or not genome size is under direct selection (Oliver *et al.*, 2007), increased genome size does appear, through its diverse impacts on plant competitiveness, plasticity, speed of adaptation or dispersal, to be negatively related to plant 'success' whether that is considering the ability of species to become invasive (Figs 1& 2), avoid becoming rare (Vinogradov, 2003), or respond to climate change (Caplat *et al.*, 2013). Having a holistic approach to understanding the status of species is therefore important (van Kleunen & Richardson, 2007; Caplat *et al.*, 2013). Mechanisms influencing genome size, apart from polyploidy, still remain to be addressed; for example, if smaller genomes proffer adaptive advantage to plant species, is this because redundant or repetitive sequences are trimmed from the genome? Even though this study does not provide answers to these questions, the clear associations that we have uncovered and the links with putative physiological mechanisms makes the study of genome size a potentially powerful tool for conservation and evolutionary biologists.

ACKNOWLEDGEMENTS

M.J.O.P. was partly supported by a NERC postdoctoral fellowship [grant number NE/F014546/1]. We thank Chan Wai Kit for help with data compilation of endangered species and Ray Callaway for constructive comments and suggestions.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 A phylogeny of the genera included in the final analysis

Fig. S2 The mean of the absolute value of latitude for each species in the dataset

Fig. S3 The relationship of genome size with chromosome number, showing the interaction with latitude.

Table S1 Comparison of phylogenetically-informed models predicting genome size from cytogenetic traits (chromosome number and ploidy level) and latitude.

Table S2 The relationship of plant invasiveness with genome size (DNA 1C-value) and cytogenetic traits (chromosome number and ploidy level), also considering the linear and quadratic effect of latitude.

Table S3 Effect of latitude on genome size (DNA 1C-value), chromosome number and invasiveness.

Table S4 Effect of the source of data on invasive species, obtained from GLMMs including genus as a random factor.

REFERENCES

APG III. **2009**. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161**: 105–121.

Bates D, Maechler M, Bolker B. **2012**. lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0.

Te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubesová M, Pysek P. **2012**. The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany* **109**: 19–45.

Bennett MD. **1972**. Nuclear DNA content and minimum generation time in herbaceous plants. *Proceedings of the Royal Society B: Biological Sciences* **181**: 109–35.

Bennett MD. **1987**. Variation in genomic form in plants and its ecological implications. *New Phytologist* **106 (Suppl)**: 177–200.

Bennett MD, Leitch IJ. **2005**. Nuclear DNA amounts in angiosperms: progress, problems and prospects. *Annals of Botany* **95**: 45–90.

Bennett MD, Leitch IJ. **2010**. *Plant DNA C-values Database*. (Release 5.0, Dec. 2010). http://data.kew.org/cvalues/

Bennett MD, Leitch IJ. **2011**. Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. *Annals of Botany* **107**: 467–590.

Bennett MD, Leitch IJ, Hanson L. **1998**. DNA amounts in two samples of Angiosperm weeds. *Annals of Botany* **82**: 121–134.

Bennett MD, Smith JB. **1972**. The effects of polyploidy on meiotic duration and pollen development in cereal anthers. *Proceedings of the Royal Society B: Biological Sciences* **181**: 81–107.

Bennett MD, Smith JB. **1991**. Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society B: Biological Sciences* **334**: 309–345.

Caplat P, Cheptou P-O, Diez J, Guisan A, Larson BMH, Macdougall AS, Peltzer DA, Richardson DM, Shea K, van Kleunen M, *et al.* 2013. Movement, impacts and management of plant distributions in response to climate change: insights from invasions. *Oikos* **122**: 1265–1274.

Caplat P, Nathan R, Buckley YM. **2012**. Seed terminal velocity, wind turbulence, and demography drive the spread of an invasive tree in an analytical model. *Ecology* **93**: 368–377.

Cavalier-Smith T. **1982**. Skeletal DNA and the evolution of genome size. *Annual Review of Biophysics and Bioengineering* **11**: 273–302.

Chamberlain S, Boettiger C, Ram K, Barve V. **2012**. rgbif: Interface to the Global Biodiversity Information Facility API methods. R package version 0.1.5.

Comai L. **2005**. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* **6**: 836–46.

Davies TJ, Wolkovich EM, Kraft NJB, Salamin N, Allen JM, Ault TR, Betancourt JL, Bolmgren K, Cleland EE, Cook BI, *et al.* 2013. Phylogenetic conservatism in plant phenology. *Journal of Ecology* **101**: 1520–1530.

Fisher DO, Owens IPF. **2004**. The comparative method in conservation biology. *Trends In Ecology & Evolution* **19**: 391–398.

Francis D, Davies MS, Barlow PW. **2008**. A strong nucleotypic effect on the cell cycle regardless of ploidy level. *Annals of Botany* **101**: 747–57.

Freckleton RP. **2009**. The seven deadly sins of comparative analysis. *Journal of Evolutionary Biology* **22**: 1367–75.

Galbraith DW, Bennetzen JL, Kellogg EA, Pires JC, Soltis PS. **2011**. The genomes of all angiosperms: a call for a coordinated global census. *Journal of Botany* **2011**: 1–10.

Gallagher R V., Leishman MR, Miller JT, Hui C, Richardson DM, Suda J, Trávníček P. **2011**. Invasiveness in introduced Australian acacias: the role of species traits and genome size. *Diversity and Distributions* **17**: 884–897.

Gregory TR. **2001**. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biological Reviews* **76**: 65–101.

Greilhuber J, Dolezel J, Lysák MA, Bennett MD. **2005**. The origin, evolution and proposed stabilization of the terms "genome size" and "C-value" to describe nuclear DNA contents. *Annals of Botany* **95**: 255–60.

Grime JP, Shacklock JML, Band SR. **1985**. Nuclear DNA contents, shoot phenology and species co-existence in a limestone grassland community. *New Phytologist* **100**: 435–445.

Hahn MA, Buckley YM, Müller-Schärer H. **2012**. Increased population growth rate in invasive polyploid Centaurea stoebe in a common garden. *Ecology Letters* **15**: 947–54.

Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretero-Paulet L, Chang T-H, Lan T, Welch AJ, Juárez MJA, Simpson J, *et al.* 2013. Architecture and evolution of a minute plant genome. *Nature* **498**: 94–98.

Ives AR, Garland T. **2010**. Phylogenetic logistic regression for binary dependent variables. *Systematic Biology* **59**: 9–26.

411 **Van Kleunen M, Dawson W, Schlaepfer D, Jeschke JM, Fischer M**. **2010a**. Are invaders
412 different? A conceptual framework of comparative approaches for assessing determinants of
413 invasiveness. *Ecology Letters* **13**: 947–58.

414 **Van Kleunen M, Richardson DM**. **2007**. Invasion biology and conservation biology: time to join
415 forces to explore the links between species traits and extinction risk and invasiveness. *Progress in*
416 *Physical Geography* **31**: 447–450.

417 **Van Kleunen M, Weber E, Fischer M**. **2010b**. A meta-analysis of trait differences between
418 invasive and non-invasive plant species. *Ecology Letters* **13**: 235–45.

419 **Knight CA, Ackerly DD**. **2002**. Variation in nuclear DNA content across environmental
420 gradients: a quantile regression analysis. *Ecology Letters* **5**: 66–76.

421 **Knight CA, Beaulieu JM**. **2008**. Genome size scaling through phenotype space. *Annals of Botany*
422 **101**: 759–66.

423 **Knight CA, Molinari NA, Petrov DA**. **2005**. The large genome constraint hypothesis: evolution,
424 ecology and phenotype. *Annals of Botany* **95**: 177–90.

425 **Kubešová M, Moravcová L, Suda J, Jarošík V, Pyšek P**. **2010**. Naturalized plants have smaller
426 genomes than their non-invading relatives: a flow cytometric analysis of the Czech alien flora.
427 *Preslia* **82**: 81–96.

428 **Kunin WE, Gaston KJ**. **1997**. *The Biology of Rarity: Causes and Consequences of Rare-*
429 *Common Differences*. London, UK: Chapman & Hall.

430 **Lavergne S, Muenke NJ, Molofsky J**. **2010**. Genome size reduction can trigger rapid phenotypic
431 evolution in invasive plants. *Annals of Botany* **105**: 109–16.

Leong-Škorničková J, Šída O, Jarolímová V, Sabu M, Fér T, Trávníček P, Suda J. **2007**. Chromosome numbers and genome size variation in Indian species of Curcuma (Zingiberaceae). *Annals of Botany* **100**: 505–26.

McIntyre PJ. **2012**. Cytogeography and genome size variation in the Claytonia perfoliata (Portulacaceae) polyploid complex. *Annals of Botany* **110**: 1195–203.

Minelli S, Moscariello P, Ceccarelli M, Cionini PG. **1996**. Nucleotype and phenotype in Vicia faba. *Heredity* **76**: 524–530.

Mowforth MA, Grime JP. **1989**. Intra-population variation in nuclear DNA amount, cell size and growth rate in Poa annua L. *Functional Ecology* **3**: 289–295.

Nakagawa S, Schielzeth H. **2013**. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4**: 133–142.

Oliver MJ, Petrov D, Ackerly D, Falkowski P, Schofield OM. **2007**. The mode and tempo of genome size evolution in eukaryotes. *Genome Research* **17**: 594–601.

Ordonez A, Wright IJ, Olff H. **2010**. Functional differences between native and alien species: a global-scale comparison. *Functional Ecology* **24**: 1353–1361.

Orgel LE, Crick FHC. **1980**. Selfish DNA: the ultimate parasite. *Nature* **284**: 604–607.

Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N. **2011**. Comparative Analyses of Phylogenetics and Evolution in R. R package version 0.4.

Pandit MK. **2006**. Continuing the search for pattern among rare plants: are diploid species more likely to be rare? *Evolutionary Ecology Research* **8**: 543–552.

Pandit MK, Pocock MJO, Kunin WE. **2011**. Ploidy influences rarity and invasiveness in plants. *Journal of Ecology* **99**: 1108–1115.

Pandit MK, Tan HTW, Bisht MS. **2006**. Polyploidy in invasive plant species of Singapore. *Botanical Journal of the Linnean Society* **151**: 395–403.

Pyšek P, Richardson DM. **2007**. Traits associated with invasiveness in alien plants: Where do we stand? In: Nentwig W, ed. Biological Invasions. Berlin: Springer, 97–125.

Pyšek P, Richardson DM. **2010**. Invasive species, environmental change and management, and health. *Annual Review of Environment and Resources* **35**: 25–55.

R Core Team. **2012**. R: A language and environment for statistical computing.

Rejmánek M. **1996**. A theory of seed plant invasiveness: The first sketch. *Biological Conservation* **78**: 171–181.

Revell LJ. **2010**. Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution* **1**: 319–329.

Schmidt JP, Drake JM. **2011**. Why are some plant genera more invasive than others? *PloS ONE* **6**: e18654.

Soltis PS, Soltis DE. **2000**. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences* **97**: 7051–7057.

Szarski H. **1983**. Cell size and the concept of wasteful and frugal evolutionary strategies. *Journal of Theoretical Biology* **105**: 201–209.

Thomson FJ, Moles AT, Auld TD, Kingsford RT. **2011**. Seed dispersal distance is more strongly correlated with plant height than with seed mass. *Journal of Ecology* **99**: 1299–1307.

Treier UA, Broennimann O, Normand S, Guisan A, Schaffner U, Steinger T, Müller-Schärer H. **2009**. Shift in cytotype frequency and niche space in the invasive plant *Centaurea maculosa*. *Ecology*: 1366–1377.

Varela-Álvarez E, Gómez Garreta A, Rull Lluch J, Salvador Soler N, Serrao EA, Siguán MAR. **2012**. Mediterranean species of *Caulerpa* are polyploid with smaller genomes in the invasive ones. *PloS ONE* **7**: e47728.

Vinogradov AE. **2003**. Selfish DNA is maladaptive: evidence from the plant Red List. *Trends in Genetics* **19**: 609–614.

Webb CO. **2000**. Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *The American Naturalist* **156**: 145–155.

Webb CO, Ackerly DD, Kembel SW. **2008**. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics* **24**: 2098–100.

Webb CO, Donoghue MJ. **2005**. Phylomatic: tree assembly for applied phylogenetics. *Molecular Ecology Notes* **5**: 181–183.

Wikström N, Savolainen V, Chase MW. **2001**. Evolution of the angiosperms: calibrating the family tree. *Proceedings Of The Royal Society B: Biological Sciences* **268**: 2211–20.

Williamson MH, Fitter A. **1996**. The characters of successful invaders. *Biological Conservation* **78**: 163–170.

Yesson C, Brewer PW, Sutton T, Caithness N, Pahwa JS, Burgess M, Gray WA, White RJ, Jones AC, Bisby FA, *et al.* 2007. How global is the global biodiversity information facility? *PloS ONE* **2**: e1124.

		Phylogenetic logistic regression (PLR)*			Generalised linear mixed model (GLMM) with genus as a random effect					
Model	Parameters	Beta	*P*	*a*†	Beta	*P*	AIC	ΔAIC ‡	$r^2_{GLMM(m)}$§	$r^2_{GLMM(c)}$§
1	Log_2 (DNA 1C-value)	-0.186	0.020	-3.02	-0.172	0.095	708.43	14.17	1.4%	15.2%
2	Log_2 (Chromosome number)	0.315	0.007	-2.71	0.519	<0.001	699.88	5.63	3.7%	17.1%
3	Log_2 (Chromosome number)	0.522	<0.001	-3.06	0.653	<0.001	694.26	0	9.0%	18.6%
	Log_2 (DNA 1C-value)	-0.311	<0.001		-0.299	0.005				
4	Log_2 (Chromosome number)	0.469	0.013	-3.02	0.609	0.006	696.18	1.92	9.3%	19.0%
	Log_2 (DNA 1C-value)	-0.440	0.326		-0.450	0.410				
	Log_2 (Chromosome number) : Log_2 (DNA C-value)	0.028	0.761		0.032	0.776				

Table 1. Effect sizes (unstandardised beta) from the relationship of plant invasiveness with genome size (DNA 1C-value; model 1) and chromosome number (model 2), both together (model 3) and together with an interaction (model 4), with the best supported model being model 3.

* We were unable to perform model selection for the PLRs due to the lack of a verified method for calculating model fit (AIC or r^2) for these types of models, so we included GLMMs to provide an assessmentof fit.

† *a* is a measure of the phylogenetic signal of the PLR; values <-2 indicate weak phylogenetic signal.

‡ΔAIC is an assessment of the relative model fit and is the difference between the model Akaike's Information Criterion (AIC) and the minimum AIC.

§r^2_{GLMM} is an assessment of the variance explained (i.e. the absolute model fit) when considering: *(m)* the fixed effects alone, and *(c)* the fixed and random effects.

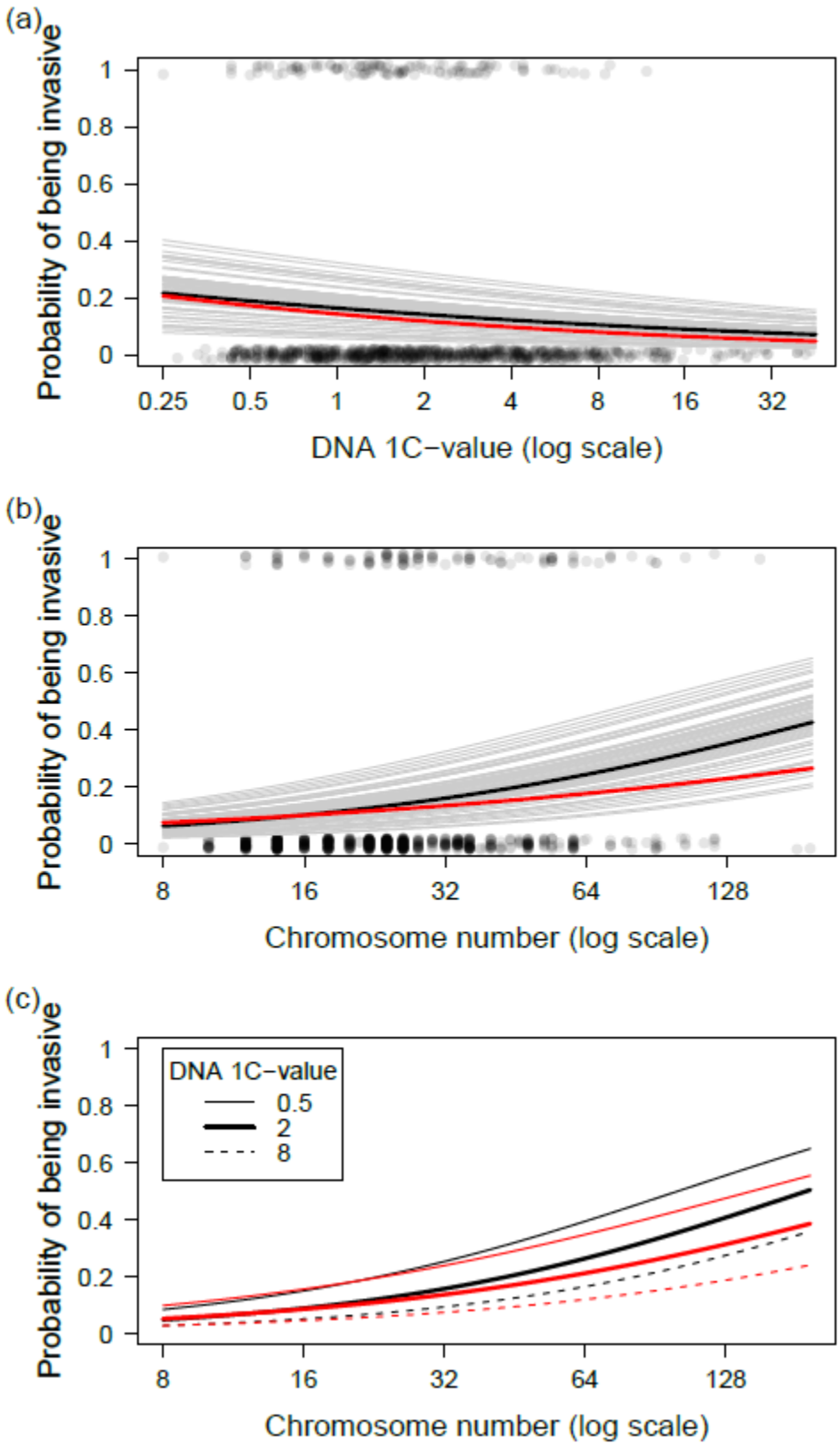
(a)
Probability of being invasive
1
0.8
0.6
0.4
0.2
0
0.25
0.5
1
2
4
8
16
32
DNA 1C−value (log scale)
(b)
Probability of being invasive
1
0.8
0.6
0.4
0.2
0
8
16
32
64
128
Chromosome number (log scale)
(c)
Probability of being invasive
DNA 1C−value
0.5
2
8
1
0.8
0.6
0.4
0.2
0
8
16
32
64
128
Chromosome number (log scale)

Fig. 1. The relationship of the probability that a species in our dataset is invasive with (a) genome size (DNA 1C-value), (b) chromosome number, and (c) chromosome number and genome size. In (a) and (b) the results of the fully phylogenetically-informed analyses (phylogenetic logistic regression; PLR) are shown in red, while from the GLMM the overall average effect is shown in black and effects for individual genera are shown in grey. Individual data points are shown as translucent points and are jittered in the y-axis for clarity. These genus-level random effects and individual data points are omitted for clarity in (c). In (c) the additive effect of genome size is presented at low, medium and high values (DNA 1C-value = 0.5, 2 and 8, respectively). Relationships with ploidy level instead of chromosome number are very similar, and so are not shown.

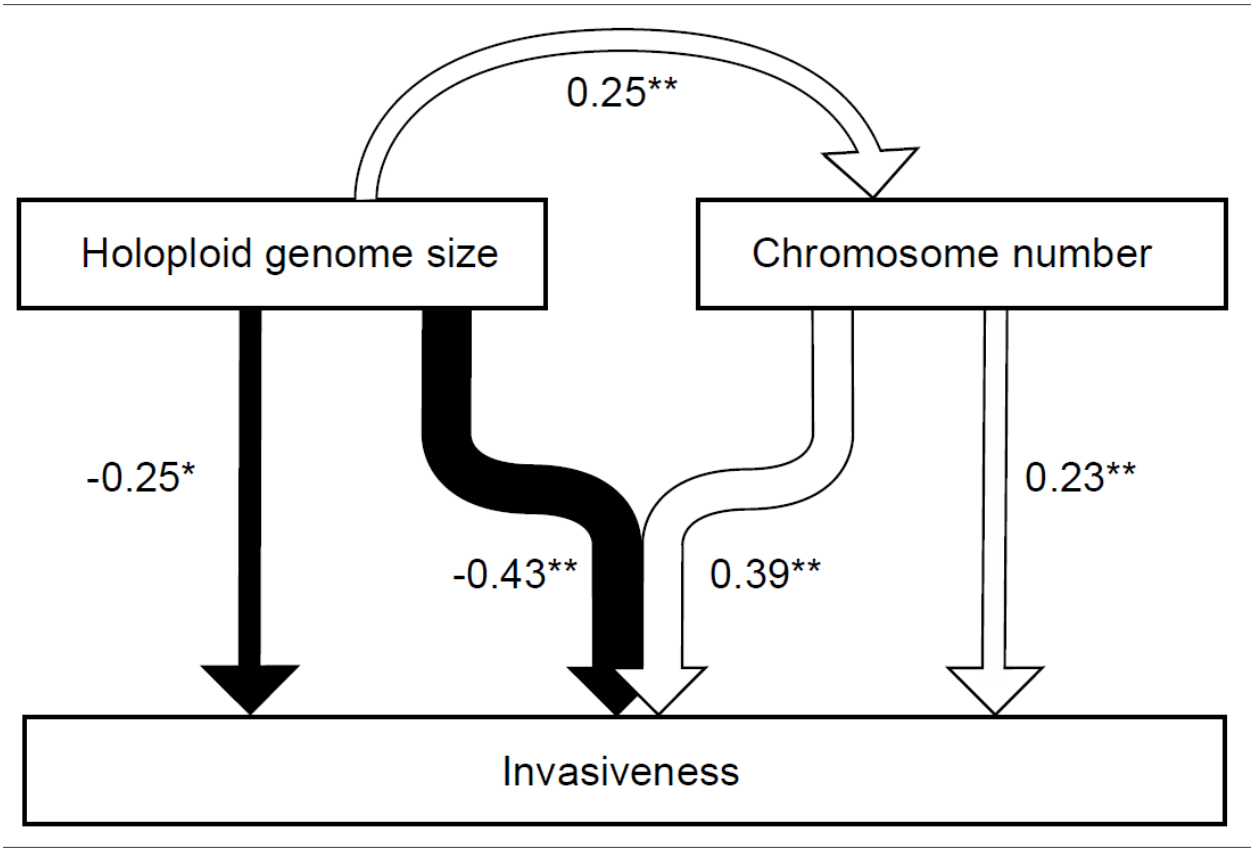

Fig. 2. Standardised effect sizes of the phylogenetic logistic regressions (PLR) between holoploid genome size (DNA 1C-value), chromosome number and invasiveness. Arrow widths are proportional to standardised effect sizes and significance is indicated by *=P<0.05 and **=P<0.001. Black arrows indicate negative relationships, white arrows indicate positive relationships. The joined arrow indicates the model in which the two traits are included as additive effects.

New Phytologist Supporting Information

Article title: The contrasting effects of genome size, chromosome number and ploidy level on plant invasiveness: a global analysis

Authors: Maharaj K. Pandit, Steven M. White and Michael J.O. Pocock

Article acceptance date: Click here to enter a date.

The following Supporting Information is available for this article:

Fig. S1 A phylogeny of the genera included in the final analysis

Fig. S2 The mean of the absolute value of latitude for each species in the dataset

Fig. S3 The relationship of genome size with chromosome number, showing the interaction with latitude.

Table S1 Comparison of phylogenetically-informed models predicting genome size from cytogenetic traits (chromosome number and ploidy level) and latitude.

Table S2 The relationship of plant invasiveness with genome size (DNA 1C-value) and cytogenetic traits (chromosome number and ploidy level), also considering the linear and quadratic effect of latitude.

Table S3 Effect of latitude on genome size (DNA 1C-value), chromosome number and invasiveness.

Table S4 Effect of the source of data on invasive species, obtained from GLMMs including genus as a random factor.

Fig. S1 A phylogeny of the genera included in the final analysis, showing the number of non-invasive ('Non') and invasive ('Inv') species included in our dataset in each genus, and the order they belong to (according to the APG III (2009)). Tree branch lengths were estimated using the 'bladj' algorithm, as described in the Methods.

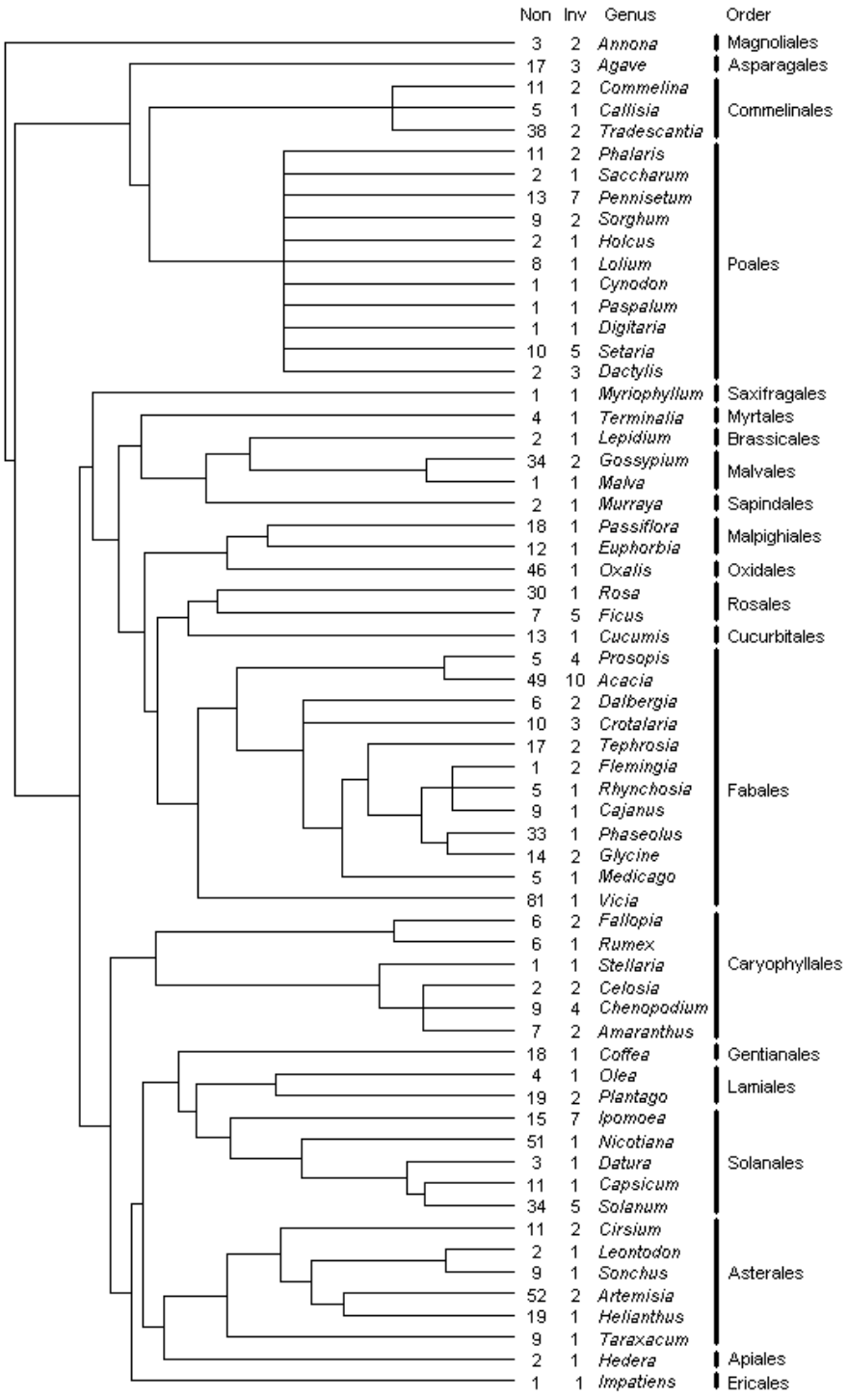
Non Inv Genus Order
3 2 Annona Magnoliales
17 3 Agave Asparagales
11 2 Commelina
5 1 Callisia Commelinales
38 2 Tradescantia
11 2 Phalaris
2 1 Saccharum
13 7 Pennisetum
9 2 Sorghum
2 1 Holcus
8 1 Lolium Poales
1 1 Cynodon
1 1 Paspalum
1 1 Digitaria
10 5 Setaria
2 3 Dactylis
1 1 Myriophyllum Saxifragales
4 1 Terminalia Myrtales
2 1 Lepidium Brassicales
34 2 Gossypium
1 1 Malva Malvales
2 1 Murraya Sapindales
18 1 Passiflora
12 1 Euphorbia Malpighiales
46 1 Oxalis Oxidales
30 1 Rosa
7 5 Ficus Rosales
13 1 Cucumis Cucurbitales
5 4 Prosopis
49 10 Acacia
6 2 Dalbergia
10 3 Crotalaria
17 2 Tephrosia
1 2 Flemingia
5 1 Rhynchosia
9 1 Cajanus
33 1 Phaseolus
14 2 Glycine
5 1 Medicago Fabales
81 1 Vicia
6 2 Fallopia
6 1 Rumex
1 1 Stellaria Caryophyllales
2 2 Celosia
9 4 Chenopodium
7 2 Amaranthus
18 1 Coffea Gentianales
4 1 Olea
19 2 Plantago Lamiales
15 7 Ipomoea
51 1 Nicotiana
3 1 Datura
11 1 Capsicum Solanales
34 5 Solanum
11 2 Cirsium
2 1 Leontodon
9 1 Sonchus
52 2 Artemisia
19 1 Helianthus Asterales
9 1 Taraxacum
2 1 Hedera Apiales
1 1 Impatiens Ericales

Fig. S2 The mean of the absolute value of latitude for each species in the dataset, as derived from records in GBIF, grouped by genus.

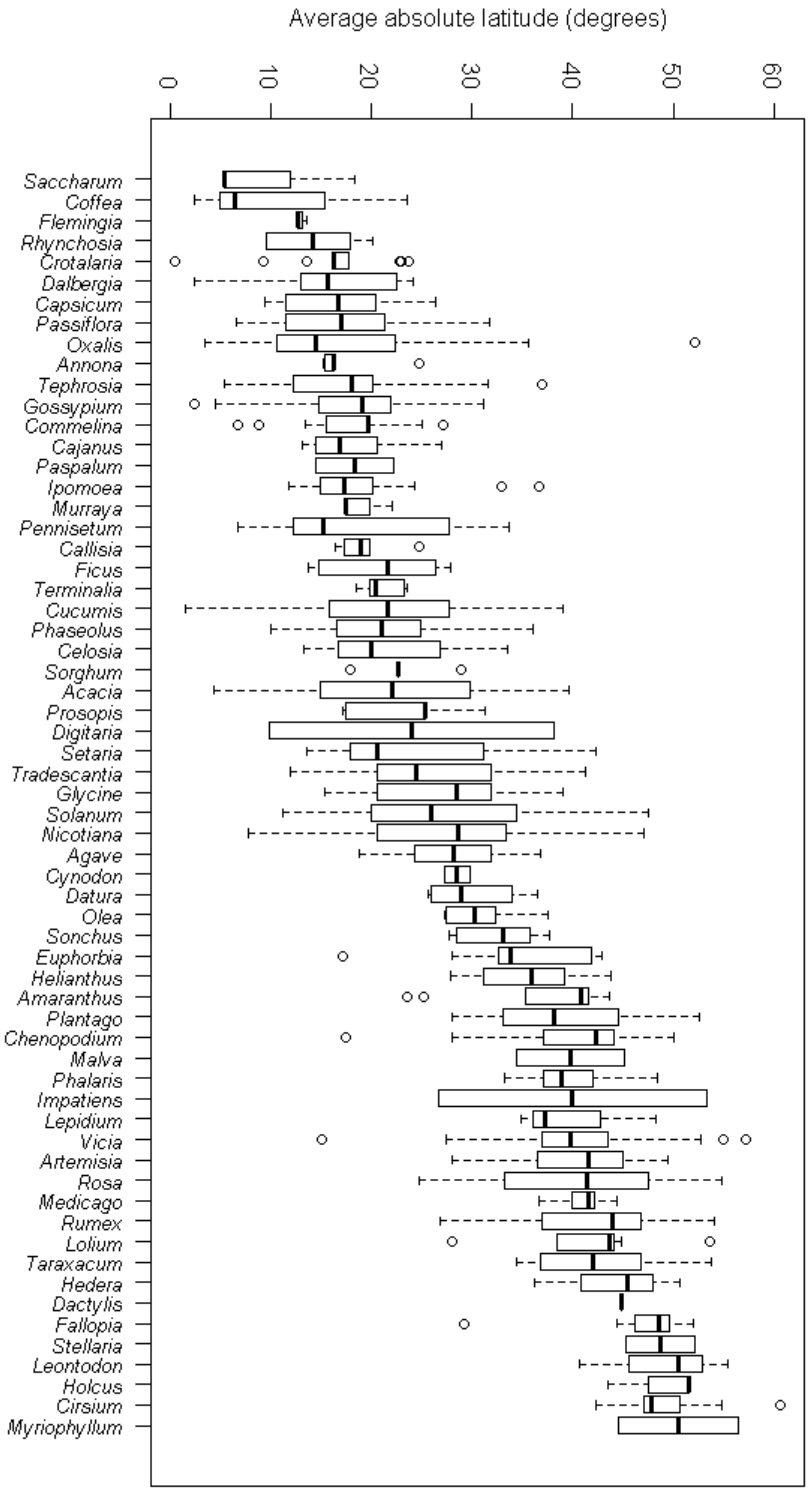

Fig. S3 The relationship of genome size (independent variable) with chromosome number from the best phylogenetic generalised least squares (PGLS) model, showing the interaction with latitude. The overall model included latitude and its interaction with genome size but for ease of interpretation, the model outputs for three reference latitudes is shown.

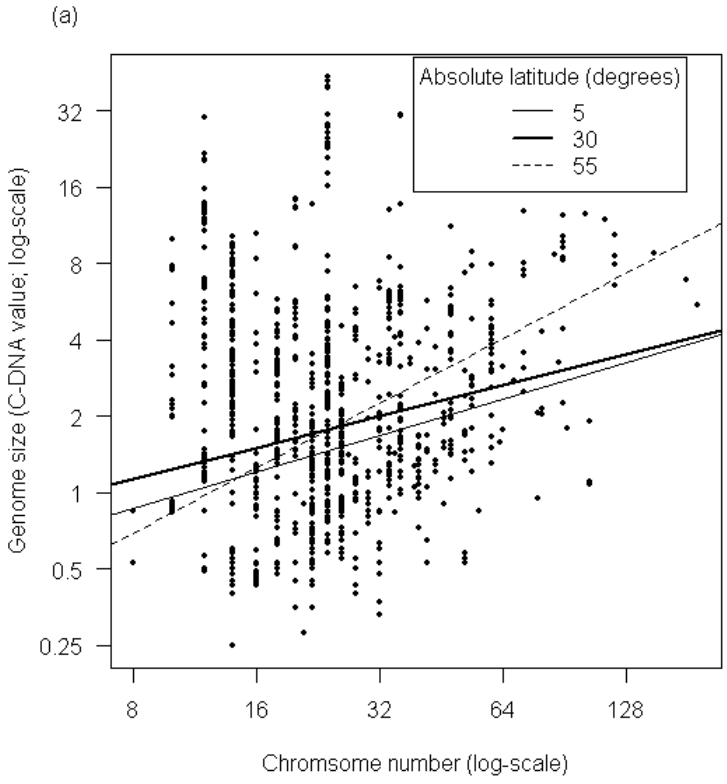

Table S1 Genome size (the dependant variable) and its relationship to cytogenetic traits (chromosome number and ploidy level) and latitude as demonstrated with phylogenetically-informed models (specifically phylogenetic generalised least squares models: PGLS). According to model selection with AIC, ploidy was a better fit to genome size than chromosome number, but for the main text we have presented our analyses with chromosome number because this is a directly-observable trait. The measure of phylogenetic signal (lambda) was close to one, indicating strong phylogenetic signal in the PGLS analyses.

		PGLS				
Model	Covariates	beta	*P*	λ	AIC	ΔAIC
1	Log2(Chromosome number)	0.460	**<0.001**	0.926	2143.23	60.18
2	Log2(Chromosome number) +	0.460	**<0.001**	0.925	2139.57	56.53
	Latitude	0.007	**0.018**			
3	Log2(Chromosome number) +	0.266	**0.005**	0.935	2136.14	53.09
	Latitude +	-0.026	0.066			
	Log2(Chromosome number): Latitude	0.007	**0.017**			
4	Log2(Chromosome number) +	0.464	**<0.001**	0.925	2138.99	55.94
	Latitude +	0.021	**0.023**			
	$Latitude^2$	-0.000	0.109			
5	Log2(Chromosome number) +	0.553	**0.001**	0.928	2134.45	51.40
	Latitude +	0.098	0.100			
	$Latitude^2$ +	-0.002	**0.034**			
	Log2(Chromosome number): Latitude +	-0.017	0.179			
	Log2(Chromosome number): $Latitude^2$	0.000	0.057			
6	Log2(Ploidy level)	0.564	**<0.001**	0.931	2094.63	11.59
7	Log2(Ploidy level) +	0.564	**<0.001**	0.931	2091.34	8.30

	Latitude	0.007	**0.022**			
8	Log2(Ploidy level) +	0.263	**0.011**	0.940	2083.04	0
	Latitude +	-0.007	0.150			
	Log2(Ploidy level): Latitude	0.010	**0.001**			
9	Log2(Ploidy level) +	0.567	**<0.001**	0.931	2090.73	7.69
	Latitude +	0.020	**0.024**			
	$Latitude^2$	-0.000	0.107			
10	Log2(Ploidy level) +	0.380	0.063	0.938	2083.60	0.56
	Latitude +	0.020	0.355			
	$Latitude^2$ +	-0.000	0.179			
	Log2(Ploidy level): Latitude +	0.000	0.977			
	Log2(Ploidy level): $Latitude^2$	0.000	0.498			

Table S2 Plant invasiveness (the dependent variable) and its relationship to genome size (DNA 1C-value) and cytogenetic traits (chromosome number and ploidy level). The linear and quadratic effect of latitude on the relationship is also shown. The dataset comprised 890 species in 62 genus from across the angiosperm phylogeny and distributed globally. Latitude was the mean of the absolute value of 1 grid cells in which the species had been recorded. 'n.c.' indicates models that did not converge. We show that the phylogenetic logistic regressions (PLR) and generalised linear mixed models (GLMM; with genus as a random effect) provide similar results. Even though it is recommended not to use multiple modelling approaches (Freckleton, 2009), we do so in order to show the similarity between the two approaches and hence justify interpretation of the AIC values, which are currently not available for PLR models.

		PLR			GLMM			
Model	Parameters	Beta	*P*	*a**	Beta	*P*	AIC	ΔAIC
With no effect of latitude								
1	Log_2(Chromosome number)	0.316	**0.007**	-2.7	0.519	**<0.001**	699.88	7.8
2	Log_2(Ploidy level)	0.372	**0.009**	-2.8	0.581	**<0.001**	700.16	8.0
3	Log_2 (DNA 1C-value)	-0.186	**0.020**	-2.8	-0.172	0.095	708.43	16.3
4	Log_2(Chromosome number)	0.522	**<0.001**	-3.1	0.653	**<0.001**	694.26	2.1
	Log_2 (DNA 1C-value)	-0.311	**0.001**		-0.299	**0.005**		
5	Log_2(Ploidy level)	0.719	**<0.001**	-3.0	0.815	**<0.001**	692.12	0.0
	Log_2 (DNA 1C-value)	-0.355	**<0.001**		-0.355	**0.001**		
6	Log_2(Chromosome number)	0.469	**0.013**	-3.0	0.609	**0.006**	696.18	4.1
	Log_2 (DNA 1C-value)	-0.440	0.326		-0.450	0.410		
	Log_2(Chromosome number): Log_2 (DNA 1C-value)	0.028	0.761		0.032	0.776		
7	Log_2(Ploidy level)	0.730	**0.001**	-3.0	0.805	**0.001**	694.12	2.0
	Log_2 (DNA 1C-value)	-0.336	0.074		-0.367	0.105		

	Log_2(Ploidy level): Log_2 (DNA 1C-value)	-0.011	0.923		0.008	0.952		
With a linear effect of latitude								
1	Log_2(Chromosome number)	0.316	**0.007**	-2.7	0.518	**<0.001**	701.78	9.7
	Latitude	-0.003	0.723		-0.003	0.752		
2	Log_2(Ploidy level)	n.c.			0.585	**<0.001**	701.88	9.8
	Latitude				-0.005	0.592		
3	Log_2 (DNA 1C-value)	-0.163	**0.035**	-4.0	-0.171	0.098	710.32	18.2
	Latitude	-0.009	0.226		-0.003	0.735		
4	Log_2(Chromosome number)	n.c.			0.652	**<0.001**	696.23	4.1
	Log_2 (DNA 1C-value)				-0.298	**0.005**		
	Latitude				-0.002	0.868		
5	Log_2(Ploidy level)	0.736	**<0.001**	-3.1	0.820	**<0.001**	693.88	1.8
	Log_2 (DNA 1C-value)	-0.353	**<0.001**		-0.354	**0.001**		
	Latitude	-0.005	0.542		-0.005	0.621		
6	Log_2(Chromosome number)	n.c.			0.608	**0.006**	698.15	6.0
	Log_2 (DNA 1C-value)				-0.448	0.413		
	Log_2(Chromosome number): Log_2 (DNA 1C-value)				0.032	0.778		
	Latitude				-0.002	0.871		
7	Log_2(Ploidy level)	0.754	**<0.001**	-3.1	0.814	**<0.001**	695.88	3.8
	Log_2 (DNA 1C-value)	-0.327	0.084		-0.360	0.113		
	Log_2(Ploidy level): Log_2 (DNA 1C-value)	-0.015	0.891		-0.005	0.623		
	Latitude	-0.005	0.502		0.004	0.977		
With a quadratic effect of latitude								

1	Log_2(Chromosome number)	0.552	**<0.001**	-2.9	0.518	**<0.001**	703.76	11.6
	Latitude	-0.007	0.832		0.003	0.935		
	Latitude^2	0.000	0.977		-0.000	0.870		
2	Log_2(Ploidy level)	0.588	**<0.001**	-3.1	0.589	**<0.001**	703.75	11.6
	Latitude	0.001	0.987		0.009	0.831		
	Latitude^2	0.000	0.702		-0.000	0.723		
3	Log_2 (DNA 1C-value)	-0.223	**0.005**	-4.0	-0.172	0.095	712.26	20.1
	Latitude	0.018	0.598		0.006	0.881		
	Latitude^2	0.000	0.469		-0.000	0.811		
4	Log_2(Chromosome number)	0.686	**<0.001**	-2.9	0.653	**<0.001**	698.12	6.0
	Log_2 (DNA 1C-value)	-0.328	**<0.001**		-0.300	**0.005**		
	Latitude	-0.014	0.678		0.011	0.781		
	Latitude^2	0.000	0.425		-0.000	0.743		
5	Log_2(Ploidy level)	1.010	**<0.001**	-3.4	0.831	**<0.001**	695.43	3.3
	Log_2 (DNA 1C-value)	-0.377	**<0.001**		-0.361	**0.001**		
	Latitude	-0.048	0.139		0.021	0.598		
	Latitude^2	0.001	0.341		-0.000	0.507		
6	Log_2(Chromosome number)	-0.156	0.426	-2.8	0.610	**0.006**	700.05	7.9
	Log_2 (DNA 1C-value)	-0.168	0.720		-0.447	0.413		
	Log_2(Chromosome number): Log_2 (DNA 1C-value)	-0.002	0.988		0.031	0.783		
	Latitude	0.001	0.978		0.011	0.784		
	Latitude^2	0.000	0.685		-0.000	0.747		
7	Log_2(Ploidy level)	1.126	**<0.001**	-3.3	0.835	**<0.001**	697.43	5.3

Log_2 (DNA 1C-value)	0.363	0.087	-0.447	0.115
Log_2(Ploidy level): Log_2 (DNA 1C-value)	-0.451	**0.002**	-0.003	0.982
Latitude	-0.042	0.200	0.021	0.597
Latitude^2	0.000	0.479	-0.000	0.507

Table S3 The effect of latitude (the independent variable) on genome size (DNA 1C-value) and chromosome number, as tested with phylogenetic generalised least squares models (PGLS), and the effect of latitude on invasiveness, as tested with phylogenetic logistic regressions (PLR). PGLS and PLR were used when the independent variable was, respectively, continuous or binary. These analyses provided different measures of phylogenetic signal: lambda which varied from 0 (no signal) to 1 (strong signal) and 'a' which varied from -4 (no signal) to +2 (strong signal). Generalised linear mixed models (GLMM) were run with genus as a random effect to provide measures of AIC for the models with invasiveness because AIC could not be calculated for PLRs. 'n.c.' indicates that the model did not converge. 'n.a.' indicates that AIC could not be calculated for PLR models. Only the effect size (beta) and significance (P) for the model covariates are shown, so '-' indicates the models with no fixed effects.

Model	Independent variable	Parameters	Phylogenetically-informed analysis					GLMM			
			Beta	*P*	AIC	ΔAIC	Phylogenetic signal*	Beta	*P*	AIC	ΔAIC
1	Log_2(DNA 1C-value)	No fixed effects	-	-	2241.32	2.89	λ = 0.84	-		2278.61	0
2		Latitude	0.0068	**0.027**	2238.44	0	λ = 0.71	0.0077	**0.014**	2284.36	5.75
3		Latitude+ Latitude^2	n.c.	n.c.	n.c.	n.c.	n.c.	0.0192 -0.0002	0.057 0.230	2300.44	21.83
1	Log_2 (Chromosome number)	No fixed effects	n.c.	n.c.	n.c.	n.c.	n.c.	-		1569.74	0
2		Latitude	-0.0007	0.763	1605.90	0.45	λ = 0.71	0.0009	0.676	1582.01	12.27
3		Latitude+	-0.0113	0.114	1605.45	0	λ = 0.72	-0.0107	0.126	1597.21	27.47

		Latitude^2	0.0002	0.117				0.0002	0.081		
1	Log_2(Ploidy level)	No fixed effects	-	-	1447.67	0	λ = 0.26	-		1573.30	0
2		Latitude	n.c.	n.c.	n.c.	n.c.	n.c.	0.0011	0.607	1585.49	12.19
3		Latitude+	-0.0075	0.252	1448.83	1.16	λ = 0.26	-0.1070	0.127	1600.61	27.30
		Latitude^2	0.0002	0.148				0.0002	0.076		
1	Invasiveness	No fixed effects	-	-	n.a.		-	-		325.06	0
2		Latitude	-0.0080	0.288	n.a.		α = -3.08	-0.0140	0.974	326.96	1.91
3		Latitude+	0.0020	0.960	n.a.		α = -3.03	-0.0150	0.995	328.94	3.89
		Latitude^2	0.00022	0.705				-0.0005	0.990		

Table S4 Effect of the source of data on invasive species, obtained from GLMMs including genus as a random factor. The datasets for GISD and PIER only was constructed exactly as described in the main text, i.e. we included all species from genera that had at least one invasive (from the specific list) and one non-invasive species. Remarkably the effect of genome size (DNA C-value) was much stronger (larger beta and smaller P value) in this analysis when considering GISD data alone, even though sample size and coverage of genera was substantially reduced, and it was comparatively more significant than the effect of chromosome number, in contrast to the other two sets of analyses.

	Invasives from GISD only	Invasives from PIER only	Invasives from either*

Model	Parameters	Beta	*P*	Beta	*P*	Beta	*P*
1	Log2 (Chromosome number)	0.439	**0.048**	0.492	**0.001**	0.519	**<0.001**
2	Log2 (DNA 1C-value)	-0.400	**0.004**	-0.191	0.066	-0.172	0.095
3	Log2 (Chromosome number)	0.560	**0.013**	0.634	**<0.001**	0.653	**<0.001**
	Log2 (DNA 1C-value)	-0.485	**0.001**	-0.313	**0.004**	-0.299	**0.005**

* repeated from Table 1 in the Main Text.