An investigation of the genome size to cell size relationship in *Equisetum.*

By
James Grayling
Supervisor Dr Barry Lomax

For decades scientists have been aware that a cell's genome size, i.e. the total amount of DNA, positively correlates with cell size. This relationship has so far been detected across all classes of vertebrate *(Gregory, 2002)* and over 100 species of Angiosperm, (flowering plants) and has been termed a 'universal phenomenon' *(Beaulieu et al, 2008).*

Since its discovery in 1951 by Mirsky and Ris, knowledge of this correlation has found application in the field of palaeogenomics, the study of ancient genomes. Here scientists use the cell size of fossilized samples as a proxy to infer the genome size (GS). *Organ et al*, (2007) traced evolution in reptiles over 200 million years using the size of fossilized bone cells to infer GS. They concluded that avian evolution occurred from reptiles with smaller genome sizes.

Yet until recently the application of GS to cell size correlation has found little use in uncovering plant GS evolution. More recent work of *Beaulieu et al (2008)* and *Knight and Beaulieu (2008)* has shed some light on the strength and ubiquity of the GS to cell size relationship in plants.

Both studies, which used over 100 species of Angiosperm, stated that GS accounts for 60-61% of guard cell length variation. Guard cells are specialized cells found either side of stomata which control gas exchange *(Smith et al, 2010)*. Knight and Beaulieu concluded that, given the strength of the relationship, there was a real possibility of inferring GS from guard cell lengths of fossilized species, to map plant GS evolution through time. Furthermore, doing so would provide the intriguing possibility of elucidating the GS response to climatic change and catastrophe.

This study will attempt to substantiate the relationship between GS and guard cell length (GCL) for 4 extant (living) species of the plant *Equisetum.* If a relationship can be confirmed, the GS of fossilized *Equisetum* can be inferred. *Equisetum* are well known for their long geological

The stem of *Equisetum arvense*.
(Image: Eric Guinther, Wikimedia Commons)

record, dating back to the upper Devonian - 360 million yr B.P *(Bateman, 1991).* If GCL measurements can be obtained from fossilized samples before, during and after events such as Angiosperm diversification and the Permo-Triassic extinction, it may be possible to quantify GS response to biotic and climatic perturbations. *(Leitch, 2007).*

While this relationship has not been confirmed for members of the Pteridophyte division (ferns) to which *Equisetum* belong, given the large-scale nature of the two studies by Knight (2008) and *Knight and Beaulieu (2008)*, as well as the similarity of their results, it is reasonable to predict that the same relationship will occur for the four species of *Equisetum* used here. A linear regression equation that reflects the relationship between GS and GCL in modern day species of *Equisetum*, can then be utilized to infer genome sizes of fossilized *Equiseta*.

Furthermore this study provides an ideal opportunity to characterize GCL with position on the plant stem. This may offer some insight as to how or whether GCL varies with plant growth. Given the primary aim of this investigation however, the less GCL variation there is (within each species) the more accurately the genome sizes of fossilized samples can be inferred. A study by *Lomax et al (2008)* has confirmed that environmental perturbations do not significantly affect the GCL of *Arabidopsis Thaliana*, increasing the integrity of the GCL to GS relationship, and thus its predictive capability. The same is assumed for this investigation.

My hypothesis and aims are therefore as follows:

Hypothesis:

1. There will be a positive correlation between guard cell length and genome size in each of the herbarium species of *Equisetum*.

2. The guard cell length of herbarium species of *Equisetum* will not vary significantly with position along the stem.

Aims:

1. To compile measurements of fossilized guard cells in order to elucidate *Equisetum*'s genome size evolution over time

2. To provide an insight as to the genome size evolution which led to *Equisetum*.

3. To provide an insight as to the environmental stresses which *Equisetum* have experienced and the resulting change in genome size and guard cell length.

Methodology:

The herbarium samples of *Equisetum* were divided into three sections per species. Three impressions per section were obtained, (nine impressions in total per species). Guard cell impressions were obtained using the standard method developed by *Coupe et al (2005)* using No.4600, President Plus, Coltene Whaledent dental putty (Altstätten, Switzerland). Each section was coated with dental putty. This was left to dry, subsequently peeled off and the impression coated with several layers of nail varnish. Once this had dried the coating was removed.

Each coating was allocated the number 1, 2 or 3, indicating the position along the stem from which it was obtained. The three coatings per section were placed on a slide in successive order obtained along the stem, and labelled with the species name. In total 10 guard cell length measurements were obtained for every section, 30 per slide, 90 per species. The guard cells were viewed under a light microscope (Zeiss Axiocam HBO $50_{/AC}$) at x40 magnification and measured using digital photography and Axio Vision 3.0 software.

Results and Discussion:

The lengths of the guard cells on the four herbarium samples of Equisetum studied here, did not change significantly with position along the stem ($P>0.05$). By excluding stem position (and assuming environmental factors do not modify GCL), it is reasonable to assume the length is a result of genetic influence. As such the predictive clarity of the linear regression equations generated here are unaffected, and so will indicate the GS of extinct *Equisetum* without bias.

Figure 1 indicates a positive correlation between GS and GCL of the four herbarium samples. This serves to reinforce the ubiquity of the relationship across the plant kingdom, with its existence now confirmed for varying forms of Angiosperms, including Moncots, Eudicots, herbs, *(Knight and Beaulieu, 2008; Beaulieu et al, 2008)*, and now for species of *Equisetum* in the Pteridophyte division. Furthermore an r^2 (correlation coefficient) value here of 0.612 is in agreement with those found by *Beaulieu et al* and *Knight* (0.62 and 0.61 respectively).

Linear regressions are usually used to find the relationship between two variables and allow prediction of *y* from *x*, using the equation $y = mx + b$. The equation in this form indicates that *x* is the independent variable and is therefore known, whereas y is the dependent variable and unknown *(Draper and Smith, 1981)*. If *y* is known however, it becomes possible to rearrange the equation and predict *x*. Thus the linear regression equations generated below to predict GS from GCL:

$$y = 0.528x + 0.9724$$

is used to predict *x* from *y*, as follows:

$$x = (y - 0.9724) / 0.528$$

As can be seen from **Fig 2**, the inferred GS values increase from the late Permian to the early Jurassic, are maintained to the late Cretaceous and fall slightly at the Eocene. While no other GS values for extinct *Equisetum* are available, prohibiting a comparison, some useful information can be extrapolated. Phenotypic correlations with GS that act against the persistence of a larger genome sizes, either through direct selection or through cellular modifications which feedback to reduce genome size, are known to exist.

Knight and Ackerly (2002) examined GS across 401 plant species in California, concluding that species with a smaller GS exist over a range of habitats, while species with larger values are more sensitive to environmental factors. This is concurrent with *Knight et al (2005)* who found that, in general, larger genomes vary much less in phenotype expression. They suggested that the pre-dominantly lower photosynthetic rates, lower Surface Leaf Area and larger seed sizes in species

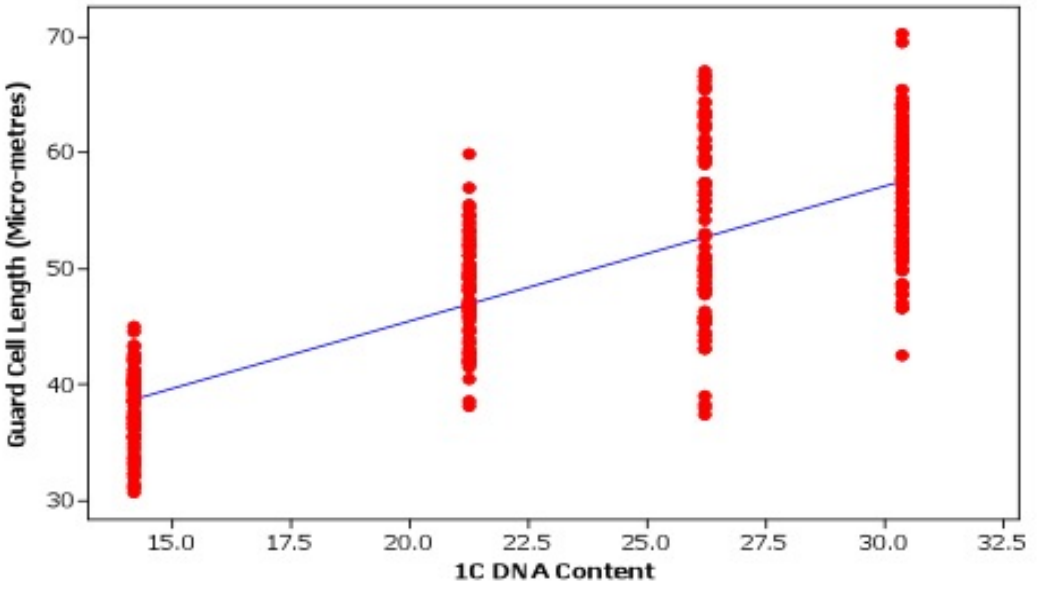

Fig. 1 *Equisetum guard cell length against genome size. The blue line is the linear regression, which log_{10} transformed = (y = 0.528x + 0.9724).*

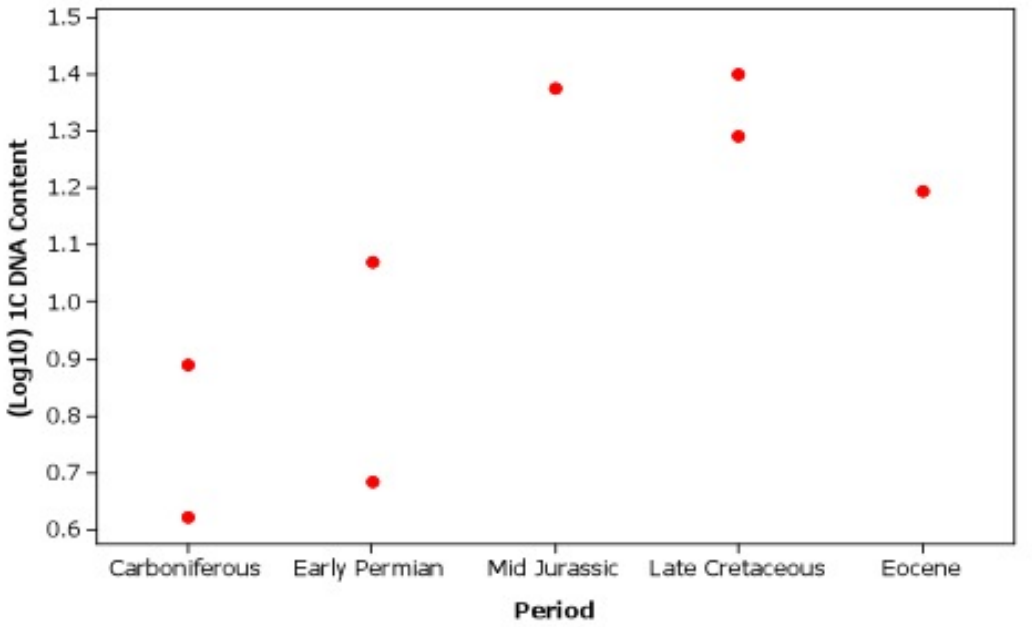

Fig. 2 *Inferred Log_{10} genome size from guard cell lengths obtained from literature plotted against their respective Periods.*

Table. 1 *Inferred genome size from guard cell lengths obtained from literature, using the Log_{10} transformed linear regression equation from Fig 1.*

Species	Period	Guard Cell Length µm	Genome size (pg)		
			x = (y - 0.9724) / 0.528		
			UPPER LIMIT		LOWER LIMIT
E.Clarnoi	Eocene	40	62.3	15.6	-37.1
E.Haukeanum	Late Cretaceous	45	71.6	19.5	-28.2
		52	93.7	25.1	-6.8
E.Filum	Mid Jurassic	50	86	23.8	-14.2
Calamitaceae	Early Permian	21.4	50.6	4.8	-47.03
		34.4	56	11.7	-42.8
Sphenophyllum	Carboniferous	20	50.4	4.17	-47.1
		28	55.7	7.76	-42.5

with larger genome sizes increase the probability of extinction.

However, there are also a plethora of genetic operations which cause GS to increase, e.g. polyploidy and transposable elements, as highlighted by *Bennetzen and Kelloggs (1997)* to support their genetic obesity hypothesis that genome sizes can only ever increase. For the purposes of this investigation, if it is assumed that the theory by *Knight et al (2005)* of a large genome constraint hypothesis is true, by identifying periods of environmental stress which act against larger genomes, it may be possible to highlight environmental factors which have influenced the GS evolution of *Equisetum*.

In reference to **Table 1**, the Annularia, which are leaves of the Genus Calamites *(Thomas and Spicer, 1987)*, were found in the peninsular of India and dated to the late early Permian. An analysis by *Stewart (1983)* concludes that of the lineages which could have given rise to *Equisetum*, including the Koretrophyllites and the Archaeocalamitaceae, it was the smaller calamitean forms which did so, 'through a reduction in size, fusion of leaves and compaction' *(Thomas and Spicer, 1987)*. If true, the ancestral GS of 4.8 pg of *Equisetaceae* is comparatively very small to modern values which range from 11.75 to 30.35 pg.

A study by *Leitch et al (1998)* concluded that the ancestral GS for Angiosperms was also small (≤1.4 - ≤ 3.5 pg), showing that the larger genome sizes were confined to specialized species that belonged to more 'derived' families *(Leitch et al 1998, Soltis et al 2003)*. They cited the small GS and therefore better reproductive rates and minimum generation time as reasons for the success of the Angiosperms in general.

Despite the advantages of a smaller GS, the Calamites were extinct by the end of the Permian, leaving behind smaller forms that crucially were capable of surviving in the drier soil *(Thomas and Spicer, 1987)*.

With reference to **Table 1**, the next inferred GS value arises in the mid Jurassic from *E. Filum Sp. Nov*. found in Yorkshire, yielding a GS value far higher than that of the Annularia of 23.8 pg. This value seems paradoxical; an increase in GS implies a lack of environmental stress, yet from the Permian to the Jurassic there were two major events, the Permo-Triassic and the Triassic-Jurassic in which changes in both climate and flora occurred.

Moving forward to late Cretaceous, for *E. Haukeanum* found in Canada, the GS values range from 19.5 to 25.1, encompassing the GS value of *E. Filum* during the Jurassic. This indicates no change and therefore no pressure on GS.

Progressing to the Eocene of the Cenozoic, *DesMarias et al (2003)* notes the beginning of the Cenozoic as a time of diversification for *Equisetum* producing the forms present today. With reference to **Table 1**, *E.Clarnoi* found in Oregon produces a slight fall in genome size to 15.6 pg. This is after the KT extinction event, now widely believed to be caused by a bollide impact *(Willis and McElwain, 2002)*. It highlights the rise of fossil ferns at the KT

boundary, with widespread wildfire providing an ideal environment for rapid re-colonization.

This is inline with *Rothwell (1996)* who highlights the rapid ability of *Equisetum* to re-colonize disturbed areas forming 'monotypic stands' following the eruption of Mt St Helens. It is possible therefore that *Equisetum* with smaller GS, capable of growing and reproducing more quickly *(Leitch et al, 1998)*, were selected during this time.

Conclusion

The primary objective of this study was to produce a linear regression equation capable of inferring the GS of fossilized *Equisetum* from GCL. This objective has been met and a positive correlation between GS and GCL in *Equisetum* has been confirmed. In addition, the GCL of Equisetum did not vary with its position along the stem, allowing the data for the four herbarium samples of *Equisetum* to be grouped and an equation produced.

The aims of this investigation were also met insofar as GCL measurements of fossilized Equisetum were collated, giving a small insight into how the genome size has evolved. In similarity with the work of *Leitch et al (2008)*, modern day *Equisetum* are derived from a small GS which, despite the advantages proposed by *Knight et al (2005)*, has subsequently increased.

Support can be offered to the genetic obesity hypothesis that GS increases only through a lack of selection against larger genome sizes *(Starostova et al, 2009)* and not as an evolutionary advantage, (as proposed by *Fawcett et al, 2009*), where a larger GS yields a higher level of adaptability. Given the changes in climate which *Equisetum* have experienced, it is reasoned that the increases in GS, rather than a lack of selection, served to increase its adaptability.

However, there were several limitations worth noting which have reduced the level of success achieved here. There were very few GCL's of *Equisetum* documented in the available literature. As such, insight into the GS evolution is at best coarse grained, with large differences in time between in each measurement. As a result only a very general analysis of environmental impacts on GS was possible, making it difficult to identify any selection pressure which may have modified the genome size. Furthermore with only one GS measurement per period, it is impossible to know whether that genome size is representative of all *Equisetaceae* at the time or just one of a multitude of other values.

In summary, despite some obstacles, this investigation provides clear evidence that the GS of *Equisetum* can be inferred from GCL measurements. Furthermore this study has demonstrated that with a larger sample more representative of the population of *Equisetum*, and when more fossilized GCL measurements become available, it will be easy to obtain an insight into GS evolution and the environmental factors which have shaped it.

Author profile:

James is 22 years of age having taken a gap year prior to University spent working and then travelling in Malaysia. James enjoyed his dissertation, taking satisfaction from contributing to a section in bioscience with currently little information. James's decision to enrol in an environmentally based degree was as a result of several enthusiastic geography teachers at secondary school. He is now working at a field centre for environmentally based studies.