



# Alliance Bioversity-CIAT Research Online

## Accepted Manuscript

### Proximal sensing of Urochloa grasses increases selection accuracy

The Alliance of Bioversity International and the International Center for Tropical Agriculture believes that open access contributes to its mission of reducing hunger and poverty, and improving human nutrition in the tropics through research aimed at increasing the eco-efficiency of agriculture.

The Alliance is committed to creating and sharing knowledge and information openly and globally. We do this through collaborative research as well as through the open sharing of our data, tools, and publications.

#### Citation:

Jiménez, Juan de la Cruz; Leiva, L.; Cardoso, J.A.; French, A.N.; Thorp, K.R. (2020) Proximal sensing of Urochloa grasses increases selection accuracy. *Crop and Pasture Science* 71(4) p. 401-409. ISSN: 1836-0947

#### Publisher's DOI:

<https://doi.org/10.1071/CP19324>

#### Access through CIAT Research Online:

<https://hdl.handle.net/10568/111686>

#### Terms:

© 2020. The Alliance has provided you with this accepted manuscript in line with Alliance's open access policy and in accordance with the Publisher's policy on self-archiving.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. You may re-use or share this manuscript as long as you acknowledge the authors by citing the version of the record listed above. You may not use this manuscript for commercial purposes. For more information, please contact Alliance Bioversity-CIAT - Library [Alliancebioversityciat-Library@cgiar.org](mailto:Alliancebioversityciat-Library@cgiar.org)

**Summary text for the online Table of Contents**

Extensive areas in the Tropics are dedicated to livestock production. Grazing activities in these areas, however, are highly restricted by forage availability. By using sensors in place of conventional methods of forage evaluation, higher number of forages can be reliably evaluated, while incurring minimal additional cost. The use of digital cameras and hyperspectral sensors to evaluate forage characteristics and production were found effective and potentially useful for selecting outstanding hybrids.

For Review Only

1 Proximal sensing of *Urochloa* grasses increases selection accuracy

2

3 Juan de la Cruz Jiménez<sup>1\*</sup>, Luisa Leiva<sup>2</sup>, Juan A. Cardoso<sup>3</sup>, Andrew N. French<sup>4</sup> and Kelly R. Thorp<sup>4</sup>

4

5 <sup>1</sup> UWA School of Agriculture and Environment, Faculty of Science, The University of Western Australia,  
6 35 Stirling Highway, Crawley, WA 6009, Australia.

7 <sup>2</sup> Department of plant breeding, Swedish University of Agricultural Sciences, Alnarp, Sweden.

8 <sup>3</sup> Tropical Forage Program, International Center for Tropical Agriculture (CIAT), Km 17 Recta Cali –  
9 Palmira, Colombia.

10 <sup>4</sup> USDA-ARS, U.S. Arid Land Agricultural Research Center, 21881 N Cardon Ln, Maricopa, AZ 85138,  
11 United States.

12

13

14

15 \* Corresponding author: Juan de la Cruz Jiménez: [juan.jimenezserna@research.uwa.edu.au](mailto:juan.jimenezserna@research.uwa.edu.au)

16

17

18

19

20

## 21 **Abstract**

22 In the American Tropics, livestock production is highly restricted by forage availability. In  
23 addition, the breeding and development of new forage varieties with outstanding yield and high  
24 nutritional quality is often limited by a lack of resources and poor technology. Non-destructive  
25 high throughput phenotyping offers a rapid and economical means to evaluate large numbers of  
26 genotypes. In this study, visual assessments, digital color images, and spectral reflectance data  
27 were collected from 200 *Urochloa* hybrids in a field setting. Partial least squares regression  
28 (PLSR) was applied to relate visual assessments, digital image analysis and spectral data with  
29 shoot dry weight (DW), crude protein (CP) and chlorophyll content. Visual evaluations of biomass  
30 and greenness, digital color imaging, and hyperspectral canopy data were collected in 68, 40 and  
31 80 minutes, respectively. Root mean squared errors of prediction for PLSR estimations of DW,  
32 CP, and chlorophyll were lower for digital image analysis followed by hyperspectral analysis and  
33 visual assessments. This study showed that digital color image and spectral analysis techniques  
34 have the potential to improve precision and reduce time for tropical forage grass phenotyping.

35 Keywords: High throughput phenotyping, *Urochloa*, tropical forage grasses, plant breeding.

36

## 37 **Introduction**

38 Livestock productivity depends on forage availability and quality. Grasses from the *Urochloa* (syn.  
39 *Brachiaria*) genus have been widely planted in the tropics as forage for grazing ruminant livestock  
40 and are considered the most important forages in the American Tropics (Miles et al. 2004). The  
41 International Center for Tropical Agriculture (CIAT) in Colombia conducts a *Urochloa* breeding  
42 program aimed at developing hybrids with outstanding performance on infertile, acidic soils with

43 superior forage productivity and nutritional quality. The hybrid development process is difficult  
44 and time consuming. In a regular, three-year breeding cycle, over 7000 hybrids are produced by  
45 open pollination, but fewer than 2% of these are retained for full evaluation. Approximately half  
46 of the population is discarded based on their reproductive mode (sexual genotypes are discarded  
47 and apomictic hybrids are kept); another major proportion is discarded based on visual evaluations;  
48 and only a limited number of hybrids (approximately 100) are finally evaluated for different biotic  
49 and abiotic stresses (Valheria Castiblanco, personal communication). The evaluation of genotypes  
50 is restricted mainly by insufficient economic resources and technology for rapid screening.

51 Forage grasses exhibiting great biomass production and high nutritional quality are key  
52 determinants of the productivity of grazing animals (Herrero et al., 2013). Therefore, evaluations  
53 of shoot biomass production and quality parameters (i.e. crude protein) are among the most  
54 important traits for improvement in any forage grasses breeding program. However, owing to the  
55 destructive nature of these measurements and the insufficient economic resources, the evaluation  
56 of these parameters is postponed to final stages of the breeding program characterized by a reduced  
57 number of genotypes. Instead of analytical measurements of forage quality and destructive  
58 biomass harvests, periodic visual evaluations of plant performance (i.e., plant biomass and  
59 greenness) over time is traditionally used in *Urochloa* breeding programs to select superior plants  
60 at initial stages of the breeding scheme (Miles et al. 2004; Miles 2007). These visual evaluations  
61 are laborious and may not be sufficiently accurate especially in breeding populations characterized  
62 by high genetic diversity and substantial genotype x environment interaction (Walter et al. 2012).  
63 The use of new technologies for in-field non-destructive, high throughput phenotyping (HTP),  
64 including digital image analysis and proximal hyperspectral sensing, offers the possibility to  
65 precisely evaluate a larger number of genotypes than feasible in traditional ways, achieved at low

66 cost, and implemented in a short period of time (Montes et al. 2007; White et al. 2012; Andrade-  
67 Sanchez et al. 2014). Proximal hyperspectral sensing provides continuous information along the  
68 visual and near-infrared electromagnetic spectrum. This information often relates to plant traits  
69 and has successfully been studied in grasses to estimate quality parameters (Skidmore et al. 2010;  
70 Pullanagari et al. 2012; Thulin et al. 2012; Ferner et al. 2015; Safari et al. 2016), diversity (Lopatin  
71 et al. 2017) and nutrient content (Fava et al. 2009; Knox et al. 2012; Ramoelo et al. 2013; Adjorlolo  
72 et al. 2015; Foster et al. 2017). Likewise, plant image analysis for phenotyping purposes is based  
73 on image segmentation to separate the soil background and the plant for further quantification of  
74 regions of interest (Tucker 1979; Woebbecke et al. 1995; Camargo 2004; Hunt et al. 2005). Digital  
75 image analysis has also been used for quantifying vegetation indices related to plant growth,  
76 greenness and nutritional status (Meyer and Camargo 2008; Hunt et al. 2013). Very few reports of  
77 hyperspectral (Numata et al. 2008) or image analysis of *Urochloa* grasses exist in literature  
78 (Jimenez et al. 2017).

79 No studies combining hyperspectral information and image analyses and comparing them to  
80 conventional phenotyping methods is available. Moreover, hyperspectral data have not been used  
81 to evaluate target traits in *Urochloa* breeding programs. In this study, in-field visual evaluations,  
82 proximal hyperspectral data, and digital imaging were collected over canopies of *Urochloa*  
83 hybrids. Partial least squares regression was used to relate hyperspectral information to field  
84 measurements and machine learning (i.e. naive Bayes multiclass) was used to extract vegetation  
85 indices from overhead canopy images. The objectives of this study were to: 1) develop PLSR  
86 models for predicting CP, forage DW, and chlorophyll content; 2) extract plant traits from digital  
87 image analysis to relate with CP, forage DW, and chlorophyll; and 3) demonstrate the superiority  
88 of HTP techniques as compared to conventional visual evaluation of traits. Crude protein, forage

89 DW, and chlorophyll content were chosen as target traits in this study as they are key parameters  
90 determining both plant and cattle productivity. The development of HTP methodologies to  
91 evaluate tropical forages will increase the number of hybrids evaluated per selection cycle, thus  
92 permitting more intense selection and hence, genetic gain. The identification of new hybrids with  
93 outstanding performance (i.e. higher biomass, greener and high CP) will result in more productive  
94 pastures with concomitant increases in milk and meat production in livestock systems in tropical  
95 savannahs.

## 96 **Materials and methods**

### 97 *Field experiment*

98 Field data were obtained in August 2016 at the International Center for Tropical Agriculture  
99 (CIAT) in Cali, Colombia (Lat. 3° 29' N; Long. 76° 21' W; altitude 965 m). Four thousand  
100 *Urochloa* hybrids generated from crosses between the CIAT's *Urochloa* breeding program  
101 population SX12 and *U. decumbens* cv. Basilisk (CIAT 606) were initially planted in an andisol  
102 soil in an augmented block design and spaced at 1.5x1.5 m. These plants were visually evaluated  
103 four times (data not shown) for persistence, vigor and greenness after sequential cuttings every  
104 three months for one year. After that period, 200 hybrids were randomly selected for further visual  
105 and HTP analysis. These 200 hybrids, instead of the entire population, were selected for economic  
106 and practical reasons. Visual evaluations of biomass and greenness, imaging and spectra collection  
107 were performed after 3 months re-growth after cutting (see information below). Plant heights  
108 ranged from 20 to 50 cm and shoot architecture varied with both decumbent and erectus growth.

### 109 *Visual evaluation*

110 Plant biomass was assessed using a nine-point visual scale, where level '9' indicated high shoot  
111 biomass with many tillers and leaves while level '1' indicated stunted growth with fewer tillers

112 and leaves. Plant greenness was visually evaluated using a five-point visual scale, where level '5'  
113 represented intense dark green in all the leaves of the plant and level '1' indicated yellow-pale  
114 color in all leaves of the plant. This visual evaluation was conducted in 68 minutes one week before  
115 the HTP measurements (Table 1).

#### 116 *Imaging collection and analysis*

117 Individual, digital color images for each of the 200 hybrids were taken at 1.2 m above the soil  
118 surface using a commercial digital 13-Megapixel camera (Coolpix P6000, Nikon, Japan) fixed to  
119 a buggy tractor. Digital images were saved in 4224 x 3168 pixel JPG format. The canopy cover  
120 (CC) and six vegetation indices including the normalized green red difference index (NGRDI),  
121 excess green index (ExG), excess red index (ExR), excess green minus excess red (ExGR), green  
122 ratio (GR) and green leaf index (GLI) were created using the formulae as indicated in Table 2. The  
123 canopy cover was extracted by dividing the total number of pixels representing the plant by the  
124 total number of pixels in each image. The vegetation indices were extracted using naive Bayes  
125 multiclass. Briefly, the distribution of colors in a set of digital color images (training set) was used  
126 to estimate the probability density function for each of the different region of interest (i.e. plant  
127 and background). Once the regions of interest were defined in the training set, the machine learning  
128 process was applied to all images to accurately classify and separate regions of interest. Therefore,  
129 every pixel in an image was classified into the previously defined plant and background classes.  
130 Every pixel characterizing the plant (but not the background) was then decomposed into red (R),  
131 green (G), and blue (B) channels. These channels were then normalized as follows:

132

$$133 \quad r = \frac{R}{R + G + B}; g = \frac{G}{R + G + B}; b = \frac{B}{R + G + B}$$

134



135 Normalization makes the variations of light intensities uniform across the spectral distribution,  
136 thus, the individual color components (i.e. r,g,b) are independent from the overall brightness of  
137 the image (Cheng et al. 2011). Normalized channels were further used for the quantification of the  
138 vegetation indices (Table 2). Image analysis code was written in Java and run in ImageJ software  
139 (National Institutes of Health, Bethesda, Maryland, USA). Images were collected early in the  
140 morning to avoid beam solar radiation interferences. Digital images contained the whole plant in  
141 addition to the 23-cm diameter field-of-view (as indicated below for hyperspectral measurements,  
142 Supplementary Fig 1). The collection process took 40 minutes (Table 1).

143

#### 144 *Spectral collection and analysis*

145 Hyperspectral field data collections were performed on clear days at full sun exposure around 11  
146 am by positioning a hand-held field spectroradiometer (Fieldspec 2, Malvern Panalytical, Malvern,  
147 UK) directly above the plant canopy. The instrument was used with no foreoptics, which provided  
148 a 25-degree full conical angle field-of-view. To avoid soil background noise, the bare optical input  
149 was positioned at 50 cm from the top of the plant canopy to yield a 23-cm diameter field of view.  
150 The instrument collected information in 750 narrow wavebands from 325 to 1075 nm in 1 nm  
151 intervals. One or ten spectral scans were collected per plant and 50 plants were evaluated daily in  
152 about 20 minutes. Differences in the collection protocols were tested to evaluate the most effective  
153 way. Different spectra collection processes (1 or 10 scans) did not yield significant differences in  
154 the root mean squared error of prediction for the different traits evaluated (Supplementary Table  
155 1). Radiometric collections over a 99% Spectralon panel (Labsphere, Inc., North Sutton, New  
156 Hampshire) were used to describe incoming solar irradiance throughout the data collection  
157 process. The radiometric collections over the calibration panel were made before starting and after  
158 every five canopy scans or when slight changes in solar irradiance due to cloud cover occurred.

159 The values of the Spectralon panel radiance were used to compute the canopy reflectance of the  
160 plants in each wavelength over the time of spectra collection. Subsequently, 401 bands from 500  
161 to 900 nm were used for analysis. Based on visual inspection of reflectance spectra, these bands  
162 were typically less noisy, as compared to bands at the bounds of detector sensitivity. Spectral  
163 collection process was run in 80 minutes (Table 1).

#### 164 *Laboratory sample collections*

165 Plants were immediately harvested after spectra collection. Aboveground tissue was removed by  
166 cutting the area defined by a 23-cm diameter plastic circle co-located with the spectral data  
167 collection area. Tissues were packed in plastic bags and stored on ice in a cooler in the field and  
168 then transported to the laboratory. The extraction of chlorophyll was performed by adding 100 mg  
169 of fresh tissue to 80% (v/v) cold methanol, and the mix was homogenized using a pestle in a mortar  
170 until the plant residue was clear and the solution was uniform. This solution was then filtered and  
171 absorbance was determined with a spectrophotometer (Synergy HT, Biotek, Winooski, USA).  
172 Total chlorophyll concentration was calculated according to Lichtenthaler and Welburn (1983).  
173 Dry weight (DW) was measured on an electronic balance (PB602S, Mettler Toledo, LLC,  
174 Columbus, OH, USA) after oven-drying the samples for three days at 60 °C. Nitrogen  
175 concentrations in the dry tissue were determined by using an automated nitrogen-carbon analyser  
176 (Sercon, Crewe, UK). *Urochloa* and common bean (*Phaseolus vulgaris*) leaves were used as  
177 reference tissues for confirmation of the reliability of the analyses. The crude protein content was  
178 calculated by multiplying nitrogen content with 6.25, as protein is assumed to contain 16%  
179 nitrogen on average.

180

#### 181 *Statistical analysis*

182 Visual evaluations, digital image analysis, spectral reflectance, and plant trait data were  
183 incorporated into a partial least squares regression (PLSR) algorithm (Mevik and Wehrens 2007)  
184 within the R Project for Statistical Computing (<http://www.r-project.org>). Models were developed  
185 to predict each plant trait (i.e. CP, DW and chlorophyll) and to compare the precision for prediction  
186 of each of the different methods of phenotyping. Partial least squares regression was used in  
187 preference to conventional least squares analysis to reduce co-linearity effects. Thorp et al. (2011)  
188 provided the details on the PLSR methodology used in the present study. Briefly, if  $\mathbf{Y}$  is an  $n \times 1$   
189 vector of responses (i.e. CP, DW or chlorophyll content) and  $\mathbf{X}$  is an  $n$ -observation by  $p$ -variable  
190 matrix of predictors (a set of visual evaluations, digital image analysis, or spectral reflectance  
191 data), PLSR aims to decompose  $\mathbf{X}$  into a set of  $A$  orthogonal scores such that the covariance with  
192 corresponding  $\mathbf{Y}$  scores is maximized. The X-weight and Y-loading vectors that result from the  
193 decomposition are used to estimate the vector of regression coefficients,  $\beta_{\text{PLS}}$ , such that

$$194 \mathbf{Y} = \mathbf{X} \beta_{\text{PLS}} + \boldsymbol{\varepsilon}$$

195 where  $\boldsymbol{\varepsilon}$  is an  $n \times 1$  vector of error terms.

196 Leave-one-out cross validation was used to test model predictions for independent data. Results  
197 were reported for PLSR models with the number of factors that minimized the root mean squared  
198 error of cross validation. Pearson's correlation coefficients were calculated for the different traits  
199 extracted from digital color images taken from *Urochloa* hybrids.

## 200 Results

201 In this study, visual evaluations of biomass and greenness, digital color imaging and hyperspectral  
202 data were collected on 200 *Urochloa* hybrids in 68, 40 or 80 minutes, respectively (Table 1). High  
203 variability for the different characteristics of DW, CP and chlorophyll content evaluated on 200  
204 *Urochloa* hybrids was found (Table 3).

205 *Visual assessments*

206 Partial least squares regressions for measured traits of DW, CP and chlorophyll based on visual  
207 evaluations of biomass and greenness performed with a root mean square error of prediction  
208 (RMSEP) of 8.47 g plant<sup>-1</sup>, 1.76% and 0.60 mg g FW respectively (Fig 1).

209 *Spectral data and digital image phenotyping*

210 The PLSR models developed from the digital image analysis estimated DW, CP and chlorophyll  
211 with a RMSEP of 7.81 g plant<sup>-1</sup>, 1.53% and 0.57 mg g FW, respectively (Fig 2). Differences on  
212 the correlation coefficients among traits extracted from image analysis indicated that including  
213 different indices into the model added independent information to build stronger PLSR models  
214 (Supplementary Fig 2). The contribution of each trait extracted from digital image analysis to the  
215 overall prediction of each destructively-measured trait is shown in Table 4. The GLI had the  
216 stronger positive influence on the PLSR model for predicting DW. The ExGR had the stronger  
217 positive influence on the PLSR model for predicting both CP and chlorophyll content.

218 The fitted PLSR models developed from 401 wavebands of canopy spectral reflectance estimated  
219 DW, CP and chlorophyll with a RMSEP of 7.90 g plant<sup>-1</sup>, 1.63% and 0.55 mg g FW, respectively  
220 (Fig 3). The contribution of each spectral waveband to the overall prediction of each destructively-  
221 measured trait is shown in the Fig 4. In the PLSR model for DW, local extrema in regression  
222 coefficients were found at 701 and 674 nm, corresponding to red light near the inflection band and  
223 red light, respectively (Fig 4a). Strong positive contribution to DW estimation were with NIR (700-  
224 750), and a strong negative contribution with red light (674-640). In the PLSR models for CP and  
225 chlorophyll, regression coefficient plots exhibited strong positive contribution for traits estimation  
226 in the visible green light (Fig 4b and c). The PLSR models for CP contrasted wavebands in the  
227 visible spectrum with positive contribution from wavebands around 503 nm and negative

228 contributions from wavebands at 678 nm. Similarly, regression coefficients for total chlorophyll  
229 indicated strong positive contribution in the visible spectrum around 504 nm and negative  
230 contribution throughout the visible wavebands, especially at 625 and 643 nm (Fig 4c). This is  
231 sensible considering visible light absorption is increased with additional leaf chlorophyll.

## 232 **Discussion**

233 The results from this study demonstrate that the current visual assessment methodology at initial  
234 steps of the breeding cycle in the CIAT *Urochloa* breeding program can be improved using non-  
235 destructive HTP techniques. Color imaging, hyperspectral analysis, and PLSR models are more  
236 precise and faster than visual evaluations, thus increasing the number of plants evaluated in the  
237 tropical forage breeding program.

238 Visual evaluations of plant growth and greenness (characteristics associated with N content, and  
239 therefore CP and chlorophyll concentration in leaves) have traditionally been used to discard  
240 *Urochloa* hybrids at initial stages of plant phenotyping. The visual evaluation of an entire breeding  
241 population (i.e., 7,000 hybrids) is a slow, costly and tedious process, and is often biased by  
242 subjectivity and human fatigue, especially when phenotypic variation of such traits is high (Table  
243 3). In this study, the estimation of DW, CP, and chlorophyll content was more precisely and  
244 consistently estimated by HTP techniques. Dry weight and CP predictions were more accurate  
245 using digital image analysis, followed by spectral analysis and visual evaluations. Chlorophyll  
246 content was better estimated by the analysis of 401 spectral wavebands, followed by color image  
247 analysis and finally visual evaluations (Fig 1, 2 and 3). The time required to run non-destructive  
248 HTP evaluations was considerably shorter by 28 minutes per 200 plants for color image analysis  
249 than visual evaluations, but longer by 12 minutes per 200 plants in hyperspectral than in visual  
250 evaluations (Table 1).

251 The moderate trends in the relationship between *Urochloa* canopy imaging and reflectance and  
252 measured DW, CP and chlorophyll may indicate that the method is not appropriate for very precise  
253 estimations of these traits. However, for breeding purposes where a large percentage of hybrids  
254 are discarded without detailed evaluation due to scarce resources, a difference in DW of 7.90 g  
255 plant<sup>-1</sup> or a difference of 1.63% in the CP content of plants may be acceptable during initial stages  
256 of plant breeding. This moderate trend between *Urochloa* canopy analysis and measured traits in  
257 this study can be explained by dissimilarities in the canopy architecture of the *Urochloa* genotypes  
258 (Numata et al. 2008), as well as different growth patterns during recovery from cutting. The further  
259 evaluation of breeding populations with contrasting canopy architecture will improve the accuracy  
260 of the PLSR model to predict the targeted traits. Nonetheless, by combining both digital image and  
261 hyperspectral analysis techniques, higher precision accuracy for DW, CP and Chlorophyll content  
262 can be achieved.

263 The vegetation indices (see Table 2) extracted from color images of 200 *Urochloa* hybrids were  
264 originally developed to separate green plants from the background by extracting green and red  
265 colors from digital images. These indices have been related to different plant characteristics  
266 including biomass, chlorophyll content and nutritional status (Tucker 1979; Woebbecke et al.  
267 1995; Camargo 2004; Hunt et al. 2005; Meyer and Camargo 2008; Hunt et al. 2013; Lee and Lee  
268 2013; Wang et al. 2013). In this study, digital image analysis performed better than hyperspectral  
269 scanning analysis to estimate DW and CP (Fig 2 and 3). Nonetheless, the use of spectral analysis  
270 over grasses becomes more important when this technique is used to detect either nutritional or  
271 anti-nutritional compounds (i.e. metabolisable energy, digestibility, fiber) that are better estimated  
272 with the near-infrared regions of the electromagnetic spectra (Curran 1989; Pullanagari et al. 2012;  
273 Ferner et al. 2015). In this sense, the use of digital color image analysis and hyperspectral analysis

274 is complementary because by using both techniques a diverse set of plant traits can accurately be  
275 predicted and by adding extra factors to the prediction model, higher prediction accuracy can be  
276 achieved (cf. Numata et al. 2008). Future efforts will use data mining to fine-tune the spectral  
277 bands included in the PLSR model (Thorp et al. 2017), which can reduce model error and improve  
278 model fit statistics. Although testing multiple methods of analysis was not the intention of this  
279 study, future research could also test other techniques (e.g., artificial neural networks) for relating  
280 HTP measurements to plant traits.

281 The regression coefficients for the PLSR for DW and chlorophyll content obtained in this study  
282 highlight that the key wavelengths for the prediction of these traits occur in the green, red, red-  
283 edge and NIR regions of the electromagnetic spectrum (Fig 4). Previous hyperspectral studies have  
284 highlighted those regions as being highly representative for dry mass and chlorophyll content in  
285 plants (Lichtenthaler et al. 1996; Thenkabail et al. 2000; Mutanga and Skidmore 2004; Fava et al.  
286 2009; Thorp et al. 2011; Adjorlolo et al. 2015; Dou et al, 2018). Although some similarities were  
287 found between wavebands among the different traits, the general regression coefficients differed  
288 among the traits, thus demonstrating that the reflectance data in a given waveband contributed  
289 differently toward the estimation of a given trait. Given the logistical burden to collect and analyze  
290 hyperspectral scans, the identification of informative key bands associated with each evaluated  
291 trait can improve the HTP process (Thorp et al. 2017). Results from this study will help guide  
292 selection of optimal bands in the construction of multispectral sensors tailored to predict specific  
293 traits of interest in tropical forage breeding programs.

294 The PLSR models for predicting DW, CP and chlorophyll content can be now used to evaluate the  
295 next generation of hybrids from the same *Urochloa* gene pool (i.e. *U. ruziziensis* – *U. brizantha* –  
296 *U. decumbens*). The accuracy of this prediction models relies on collection protocols similar to the

297 explained in the Materials and Methods section and evaluations on plants with comparable growth  
298 characteristics as the hybrids evaluated here (i.e. about three months after regrowth). The  
299 prediction accuracy will likely be reduced on larger plants with higher biomass (Hill 2004) and a  
300 greater proportion of senescent leaves (Asner 1998). The development of more precise PLSR  
301 models to predict variables of interest in a breeding program requires an ongoing effort. The  
302 collection of ground data every year while making improvements to standardize collection  
303 protocols and incorporate wider range of genotypes will result in more accurate and robust models.  
304 Larger data sets will increase estimation precision.

305

## 306 **Conclusions**

307 In this study, 200 *Urochloa* hybrids were monitored in 40 and 80 minutes by digital imaging and  
308 spectral analysis, respectively (Table 1). At this pace, more than 1000 *Urochloa* hybrids could be  
309 evaluated in a period of less than 7 hours. This means that forage biomass and quality in a high  
310 number of genotypes would be reliably evaluated with minimal increased acquisition costs relative  
311 to destructive harvest. This demonstrates the superiority of HTP techniques as compared to  
312 conventional visual evaluation of traits. The PLSR models for predicting CP, forage DW, and  
313 chlorophyll content developed in this study supports the evaluation of higher numbers of genotypes  
314 at initial stages of the breeding program. The greater numbers of plants evaluated reliably every  
315 year in the *Urochloa* breeding program, the greater the genetic gain will be. Therefore, the use of  
316 image analysis and hyperspectral monitoring over *Urochloa* hybrids canopies will benefit the on-  
317 going breeding program. The application of this HTP method could be of great help in rural remote  
318 areas lacking facilities to perform destructive harvest and plant chemical analysis. Research is  
319 underway to improve the utility of proximal sensing by considering a greater range of canopy



320 architectural configurations and evaluating the potential to assess nutritional quality, including  
321 characteristics such as metabolisable energy, fiber, digestibility, lignin and cellulose fractions in  
322 *Urochloa* grasses.

323

#### 324 **Conflict of interest**

325 The authors have no conflicts of interest to declare.

#### 326 **Acknowledgements**

327 We thank Dr. John W. Miles for helpful suggestions and comments to early version of this  
328 manuscript. JCJ is grateful to the USDA Foreign Agricultural Service, Norman Borlaug  
329 Fellowship Program for a training fellowship. JCJ thanks the CIAT's Young Scientist Award 2016  
330 Program for travel assistance. This work was partially undertaken as part of the CGIAR Research  
331 Program on Livestock. We thank all donors that globally support our work through their  
332 contributions to the CGIAR system.

#### 333 **References**

334 Adjorlolo C, Mutanga O, Cho MA (2015) Predicting C3 and C4 grass nutrient variability using *in*  
335 *situ* canopy reflectance and partial least squares regression. *International Journal of Remote*  
336 *Sensing* **36(6)**, 1743–1761.

337

338 Andrade-Sanchez P, Gore MA, Heun JT, Thorp KR, Carmo-Silva AE, French AN, Salvucci ME,  
339 White JW (2014) Development and evaluation of a field-based high throughput phenotyping  
340 platform. *Functional Plant Biology* **41**, 68–79.

341

- 342 Camargo JN (2004) A combined statistical—soft computing approach for classification and  
343 mapping weed species in minimum tillage systems. University of Nebraska, Lincoln, NE  
344
- 345 Carrant PJ (1989) Remote sensing of foliar chemistry. *Remote Sensing of the Environment* **30**,  
346 271–278.  
347
- 348 Dou Z, Cui L, Li J, Zhu Y, Gao C, Pan X, Lei Y, Zhang M, Zhao X, Li W (2018) Hyperspectral  
349 estimation of the chlorophyll content in short-term and long-term restorations of mangrove in  
350 Quanzhou Bay Estuary, China. *Sustainability* **10**, 1127; doi:10.3390/su10041127  
351
- 352 Fava F, Colombo R, Bocchi S, Meroni M, Sitzia M, Fois N, Zucca C (2009) Identification of  
353 hyperspectral vegetation indices for Mediterranean pasture characterization. *International Journal*  
354 *of Applied Earth Observation and Geoinformation* **11**, 233–243.  
355
- 356 Ferner J, Linstadter A, Sudekum KH, Schmidlein S (2015) Spectral indicators of forage quality  
357 in West Africa's tropical savannas. *International Journal of Applied Earth Observation and*  
358 *Geoinformation* **41**, 99–106.  
359
- 360 Foster AJ, Kakani VG, Mosali J (2017) Estimation of bioenergy crop yield and N status by  
361 hyperspectral canopy reflectance and partial least square regression. *Precision Agriculture* **18**,  
362 192–209.  
363
- 364 Herrero M, Havlik P, Valin H, Notenbaert A, Rufino MC, Thornton PK, Blümmel M, Weiss F,  
365 Grace D, Obersteiner M (2013) Biomass use, production, feed efficiencies, and greenhouse gas

366 emissions from global livestock systems. *Proceedings of the National Academy of Science of the*  
367 *United States of America* **110(52)**, 20888–20893.

368

369 Hunt Jr ER, Cavigelli M, Daughtry CST, McMurtrey J, Walthall CL (2005) Evaluation of digital  
370 photography from model aircraft for remote sensing of crop biomass and nitrogen status. *Precision*  
371 *Agriculture* **6**, 359–378.

372

373 Hunt Jr ER, Doraiswamy PC, McMurtrey J, Daughtry CST, Perry EM, Akhmedov B (2013) A  
374 visible band index for remote sensing leaf chlorophyll content at the canopy scale. *International*  
375 *Journal of Applied Earth Observation and Geoinformation* **21**, 103–112.

376

377 Jiménez JC, Cardoso JA, Leiva LF, Gil J, Forero MG, Worthington ML, Miles JW, Rao IM (2017)  
378 Non-destructive phenotyping to identify brachiaria hybrids tolerant to waterlogging stress under  
379 field conditions. *Frontiers in Plant Science* **8**, 167. doi: 10.3389/fpls.2017.00167

380

381 Knox NM, Skidmore AK, Prins HHT, Heitkönig IMA, Slotow R, van der Waal C, de Boer WF  
382 (2012) Remote sensing of forage nutrients: Combining ecological and spectral absorption feature  
383 data. *ISPRS Journal of Photogrammetry and Remote Sensing* **72**, 27–35.

384

385 Meyer GE, Hindman TW, Lakshmi K (1998) Machine vision detection parameters for plant  
386 species identification. In 'Precision Agriculture and Biological Quality'. (Eds GE Meyer, JA De  
387 Shazer) pp. 327–335. (Proceedings of SPIE. vol. 3543, Bellingham, WA)

388

389 Meyer GE, Camargo J (2008) Verification of color vegetation indices for automated crop imaging  
390 applications. *Computers and Electronics in Agriculture* **63**, 282–293.

391

392 Mevik BH, Wehrens R (2007) The pls Package: Principal component and partial least squares  
393 regression in R. *Journal of Statistical Software* **18(2)**, 1-23.

394

395 Miles JW, Do Valle CB, Rao IM, Euclides VPB (2004) Brachiariagrasses. American Society of  
396 Agronomy, Crop Science Society of America, Soil Science Society of America, 677 S. Segoe Rd.,  
397 Madison, WI 53711, USA. *Warm Season (C4) Grasses*, Agronomy Monograph no. 45.

398

399 Miles JW (2007) Apomixis for cultivar development in tropical forage grasses. *Crop Science*  
400 **47(S3)**, S238–S249.

401

402 Montes JM, Melchinger AE, Reif JC (2007) Novel throughput phenotyping platforms in plant  
403 genetic studies. *Trends in Plant Science* **12**, 433–436.

404

405 Mutanga O, Skidmore AK (2004) Narrow band vegetation indices overcome the saturation  
406 problem in biomass estimation. *International Journal of Remote Sensing* **25(19)**, 3999–4014.

407

408 Numata I, Roberts DA, Chadwick OA, Schimel JP, Galvão LS, Soares JV (2008) Evaluation of  
409 hyperspectral data for pasture estimate in the Brazilian Amazon using field and imaging  
410 spectrometers. *Remote Sensing of Environment* **112**, 1569–1583.

411

- 412 Lee KJ, Lee BW (2013) Estimation of rice growth and nitrogen nutrition status using color digital  
413 camera image analysis. *European Journal of Agronomy* **48**, 57–65.
- 414
- 415 Lichtenthaler HK, Wellburn AA (1983) Determination of total carotenoids and chlorophylls a and  
416 b of leaf extracts in different solvents. *Biochemical society transactions* **603**, 591–592.
- 417
- 418 Lichtenthaler HK, Gitelson A, Lang M (1996) Non-destructive determination of chlorophyll  
419 content of leaves of a green and an aurea mutant of tobacco by reflectance measurements. *Journal*  
420 *of Plant Physiology* **148**, 483–493.
- 421
- 422 Lopatin J, Fassnacht FE, Kattenborn T, Schmidlein S (2017) Mapping plant species in mixed  
423 grassland communities using close range imaging spectroscopy. *Remote Sensing of Environment*  
424 **201**, 12–23.
- 425
- 426 Louhaichi M, Borman MM, Johnson DE (2001) Spatially located platform and aerial photography  
427 for documentation of grazing impacts on wheat. *Geocarto International* **16(1)**, 65-70.
- 428
- 429 Pullanagari RR, Yule IJ, Tuohy MP, Hedley MJ, Dynes RA, King WM (2012) In-field  
430 hyperspectral proximal sensing for estimating quality parameters of mixed pasture. *Precision in*  
431 *Agriculture* **13**, 351–369.
- 432
- 433 Ramoelo A, Skidmore AK, Schlerf M, Heitkönig IWA, Mathieu R, Cho MS (2013) Savanna grass  
434 nitrogen to phosphorous ratio estimation using field spectroscopy and the potential for estimation

435 with imaging spectroscopy. *International Journal of Applied Earth Observation and*  
436 *Geoinformation* **23**, 334–343.

437

438 Safari H, Fricke T, Wachendorf M (2016) Determination of fibre and protein content in  
439 heterogeneous pastures using field spectroscopy and ultrasonic sward height measurements.  
440 *Computers and Electronics in Agriculture* **123**, 256–263.

441

442 Skidmore AK, Ferwerda JG, Mutanga O, Van Wieren SE, Peel M, Grant RC, Prins HHT, Balcik  
443 FB, Venus V (2010) Forage quality of savannas - Simultaneously mapping foliar protein and  
444 polyphenols for trees and grass using hyperspectral imagery. *Remote Sensing of Environment* **114**,  
445 64–72.

446

447 Thenkabail PS, Smith RB, Pauw ED (2000) Hyperspectral vegetation indices and their  
448 relationships with agricultural crop characteristics. *Remote Sensing of Environment* **71**, 158–182.

449

450 Thorp KR, Dierig DA, French AN, Hunsaker DJ (2011) Analysis of hyperspectral reflectance data  
451 for monitoring growth and development of lesquerella. *Industrial Crops and Products* **33**, 524–  
452 531.

453

454 Thorp KR, Wang G, Bronson KF, Badaruddin M, Mon J (2017) Hyperspectral data mining to  
455 identify relevant canopy spectral features for estimating durum wheat growth, nitrogen status, and  
456 grain yield. *Computers and Electronics in Agriculture* **136**, 1–12.

457

458 Thulin S, Hill MJ, Held A, Jones S, Woodgate P (2012) Hyperspectral determination of feed  
459 quality constituents in temperate pastures: Effect of processing methods on predictive relationships  
460 from partial least squares regression. *International Journal of Applied Earth Observation and*  
461 *Geoinformation* **19**, 322–334.

462

463 Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation.  
464 *Remote sensing of environment* **8**, 127-150.

465

466 Walter A, Studer B, Kölliker R (2012) Advanced phenotyping offers opportunities for improved  
467 breeding of forage and turf species. *Annals of Botany* **110**, 1271–1279.

468

469 Wang Y, Wang D, Zhang G, Wang J (2013) Estimating nitrogen status of rice using the image  
470 segmentation of G-R thresholding method. *Field Crops Research* **149**, 33–39.

471

472 Woebecke DM, Meyer GE, Von Bargen K, Mortensen DA (1995) Color indices for weed  
473 identification under various soil, residue, and lighting conditions. *Transactions of the ASAE* **38(1)**,  
474 259-269.

475

476

477

478

479

480

481

482 **Table 1.** Phenotyping techniques used in the present study, the time of evaluation, its application,  
 483 advantages and disadvantages.

484

Phenotyping technique	Time of evaluation*	Applications	Advantages	Disadvantages
<b>Visual evaluation</b>	68 min	Visual observations of different plant characteristics	Easy operation, low cost, evaluations can be performed under diverse conditions and environments	Evaluation of low number of genotypes, evaluation subjected to human bias and fatigue
<b>Image analysis</b>	40 min	Quantification of canopy cover and vegetation indices in the visible electromagnetic spectrum	Easy operation, low cost, greater number of plants evaluated, determination of several vegetation and water indices	Changes in ambient light conditions limit calculation of vegetation indices, data analysis is moderately complex
<b>Hyperspectral analysis</b>	80 min	Canopy reflectance information in the visible and near infra-red regions of the electromagnetic spectrum. Information can be used to predict biochemical composition of plants	Moderately easy operation, greater number of plants evaluated, determination of nutritional and biochemical composition of leaf/canopy	Low solar radiation or cloudy days limit analysis, sensor and white reference calibration is frequently needed, data analysis is complex

485 \* The time of evaluation refers to 200 *Urochloa* plants evaluated under the conditions of the  
 486 present study.

487

488

489

490

491

492

493

494



495 **Table 2.** Canopy cover and vegetation indices calculated from digital images of 200 *Urochloa*  
 496 hybrids. Vegetation indices were extracted using a naive Bayes multiclass machine learning  
 497 approach. Indices were then incorporated into a PLSR model to predict crude protein, dry weight  
 498 biomass and chlorophyll content.  
 499

Plant traits	Name	Formula*	Reference
<b>CC**</b>	Canopy cover	$N_c/N_t$	-
<b>NGRDI</b>	Normalized green red difference index	$(g-r)/(g+r)$	Hunt et al., 2005
<b>ExG</b>	Excess green index	$2g-r-b$	Woebbecke et al., 1995
<b>ExR</b>	Excess red index	$1.3r-g$	Meyer et al., 1998
<b>ExGR</b>	Excess green minus excess red	$ExG-ExR$	Camargo 2004
<b>GR</b>	Green ratio	$g/(r+g+b)$	Tucker 1979
<b>GLI</b>	Green leaf index	$(2g-r-b)/(2g+r+b)$	Louhaichi et al. 2001

\*r, g and b denote the normalized pixel values of each channel on the RGB colour mode.

\*\* No normalization was performed for the canopy cover quantification.  $N_c$ = total number of pixels representing the canopy,  $N_t$ = total number of pixels in the picture.

500

501

502

503

504 **Table 3.** Plant traits measured in 200 *Urochloa* hybrids.

505

<b>Trait</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>CV (%)</b>
<b>Dry Weight (g plant<sup>-1</sup>)</b>	6.74	64.1	30.22	34.81
<b>Crude Protein (%)</b>	6.76	21.58	11.23	19.68
<b>Chlorophyll (mg g FW)</b>	0.87	6.41	2.88	24.31

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

For Review Only

521 **Table 4.** Regression coefficients of the fitted partial least square regression models of seven traits  
 522 extracted from digital image analysis. Positive and negative coefficients indicate positive and  
 523 negative influence on the prediction model, respectively.

<b>Traits*</b>	<b>Dry weight (g.plant<sup>-1</sup>)</b>	<b>Crude protein (% DW)</b>	<b>Chlorophyll (mg.g<sup>-1</sup>)</b>
<b>CC</b>	3.760545	-0.23114345	-0.01673706
<b>NGRDI</b>	9.948634	0.08600517	-0.03532585
<b>ExG</b>	-14.3163	0.07760486	0.0730642
<b>ExR</b>	-32.126212	-0.39106455	-0.07758547
<b>ExGR</b>	3.724492	0.26592971	0.10326555
<b>GR</b>	-34.770799	-0.31158671	-0.03624834
<b>GLI</b>	80.87152	-0.31108316	-0.0357783

524 \* CC= canopy cover, NGRDI= normalized green red difference index, ExG= excess green index,  
 525 ExR= excess red index, ExGR= excess green minus excess red, GR= green ratio and GLI= green  
 526 leaf index.  
 527

528

529

530

531

532

533

534

535

536

537

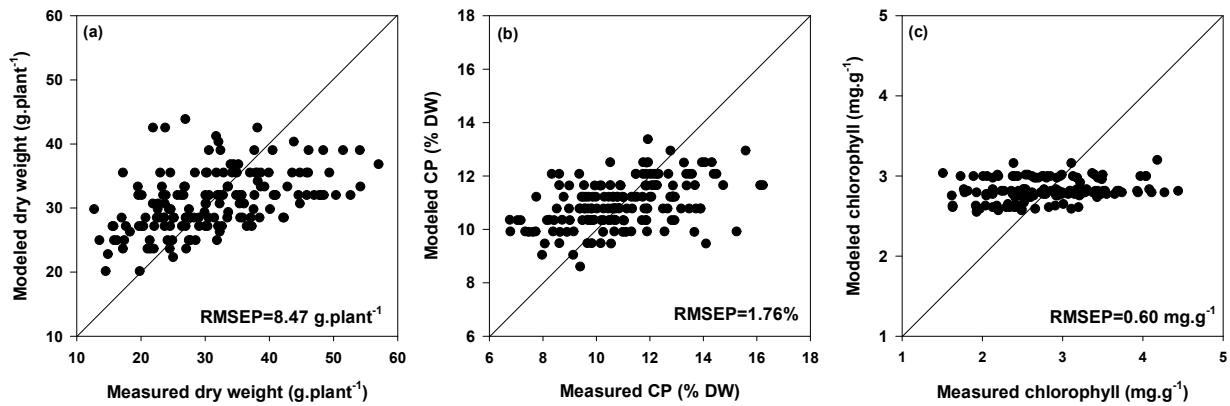
538

539

540 **Fig 1.** Modeled versus measured dry weight, crude protein and chlorophyll content when fitting  
541 partial least square regression models to relate each biophysical characteristic to visual evaluations  
542 of biomass and greenness of 200 *Urochloa* hybrids.

543

544



545

546

547

548

549

550

551

552

553

554

555

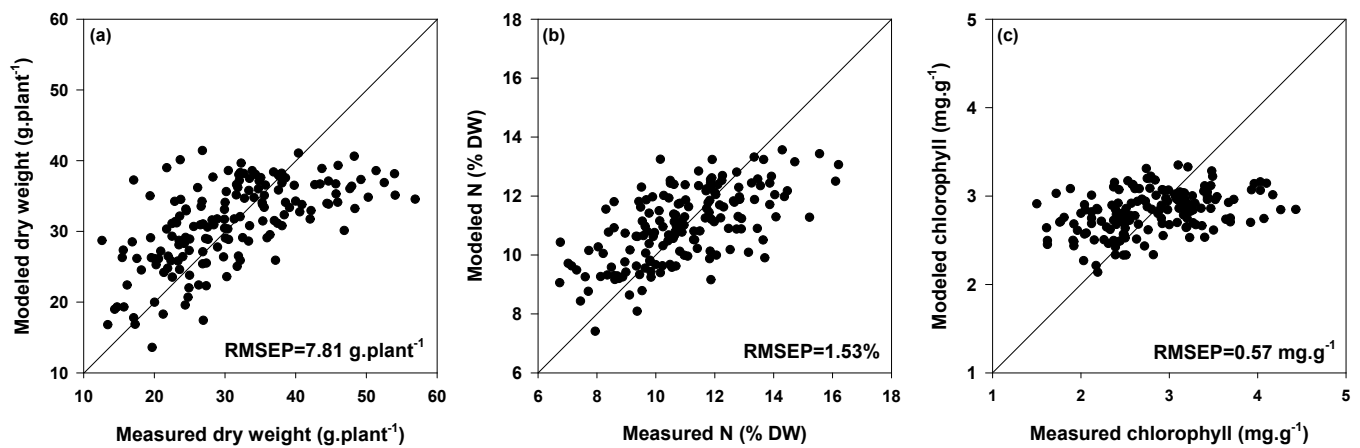
556

557 **Fig 2.** Modeled versus measured dry weight, crude protein and chlorophyll content when fitting  
558 partial least square regression models to relate each biophysical characteristic to digital image  
559 analysis of 200 *Urochloa* hybrids.

560

561

562



563

564

565

566

567

568

569

570

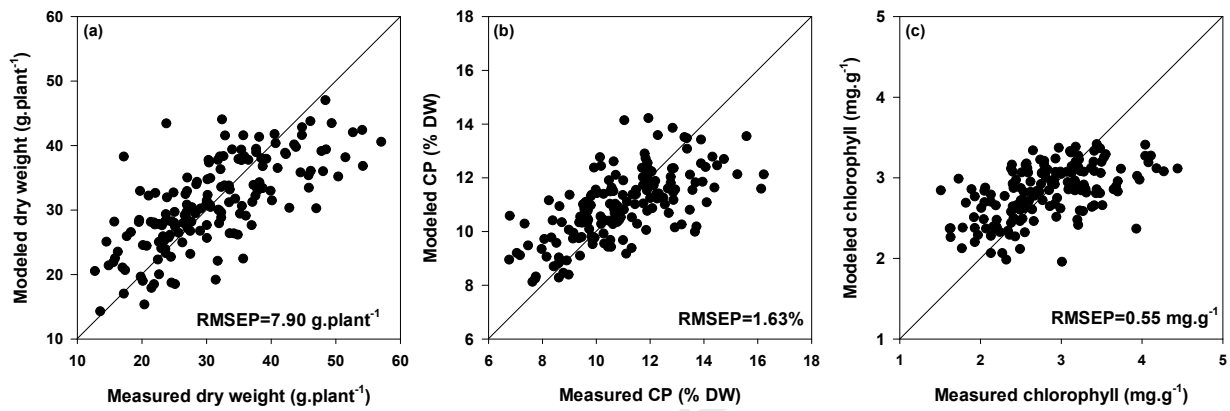
571

572 **Fig 3.** Modeled versus measured dry weight, crude protein and chlorophyll content when fitting  
573 partial least square regression models to relate each biophysical characteristic to canopy spectral  
574 reflectance of 200 *Urochloa* hybrids.

575

576

577



578

579

580

581

582

583

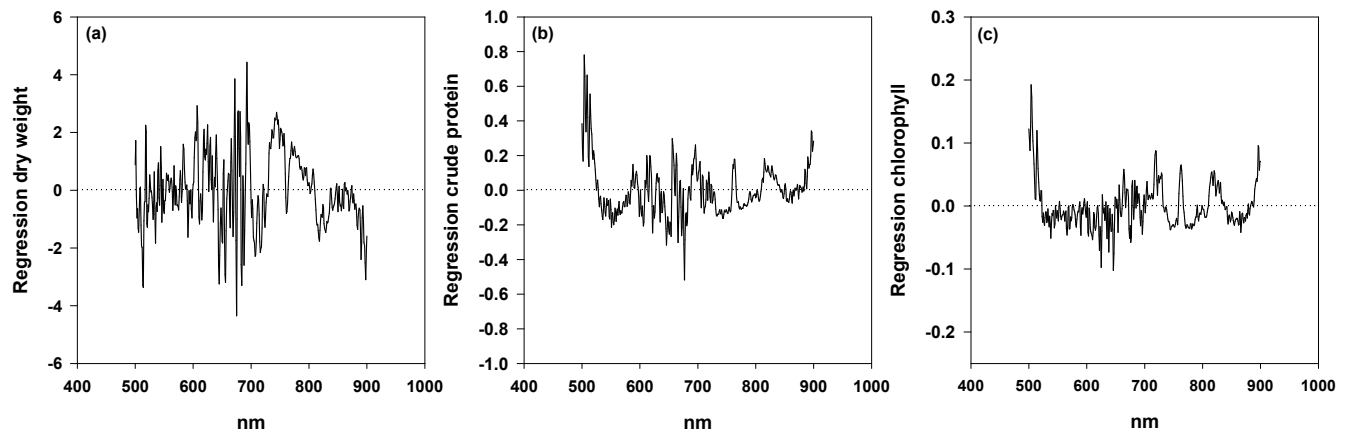
584

585

586

587 **Fig 4.** Regression coefficients of the fitted partial least squares regression models for dry weight,  
588 crude protein and chlorophyll content. The regression coefficients represents the contribution of  
589 each spectral waveband to the overall prediction of each destructively-measured trait.

590



591

592

593

594

595

596

597

598

599

600

601

602

603

604 **Supplementary information**

605 **Supplementary Table 1.** Different protocols of spectral data collection and their respective root  
 606 mean squared error of prediction (RMSEP) for crude protein, dry weight and chlorophyll content.

607	Collection	Trait	Factors¥	RMSEP
608	Day 1*	Dry weight (g plant <sup>-1</sup> )	4	9.23
		Crude protein (%)	11	1.29
		Chlorophyll (mg g FW)	4	0.49
609	Day 2*	Dry weight (g plant <sup>-1</sup> )	4	8.20
		Crude protein (%)	10	1.26
		Chlorophyll (mg g FW)	7	0.50
610	Day 3**	Dry weight (g plant <sup>-1</sup> )	4	7.63
		Crude protein (%)	11	2.07
		Chlorophyll (mg g FW)	5	0.54
611	Day 4**	Dry weight (g plant <sup>-1</sup> )	6	8.14
		Crude protein n (%)	2	1.21
612		Chlorophyll (mg g FW)	3	0.58
613	All days	Dry weight (g plant <sup>-1</sup> )	6	7.90
		Crude protein (%)	5	1.63
		Chlorophyll (mg g FW)	5	0.55

614 Fifty plants were evaluated daily \* One scan collected per plant. \*\* Ten scans collected per plant.  
 615 ¥ Number of factors for which the root mean squared error of prediction was minimized in the  
 616 model prediction.

617

618

619

620

621

622

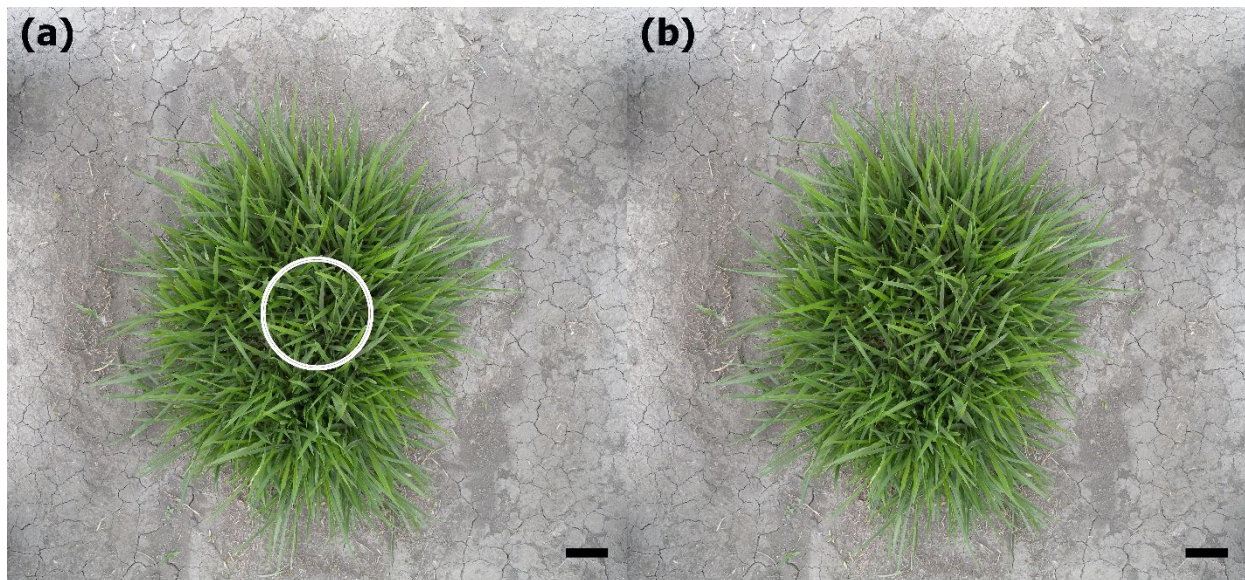
623

624

625



626 **Supplementary Fig 1.** Schematic representation of the observation geometry of hyperspectral  
627 analysis (a) and digital image analysis (b) techniques evaluated in 200 *Urochloa* hybrids. White  
628 circle positioned at the center of the plant canopy in figure (a) represents the 23-cm field of view  
629 of the spectroradiometer at a distance of 50mm from the plant canopy. For the digital image  
630 analysis (figure b), the whole plant, and not the 23-cm section, was used for segmentation and  
631 further analysis. Scale bar= 10 cm.



632

633

634

635

636

637

638

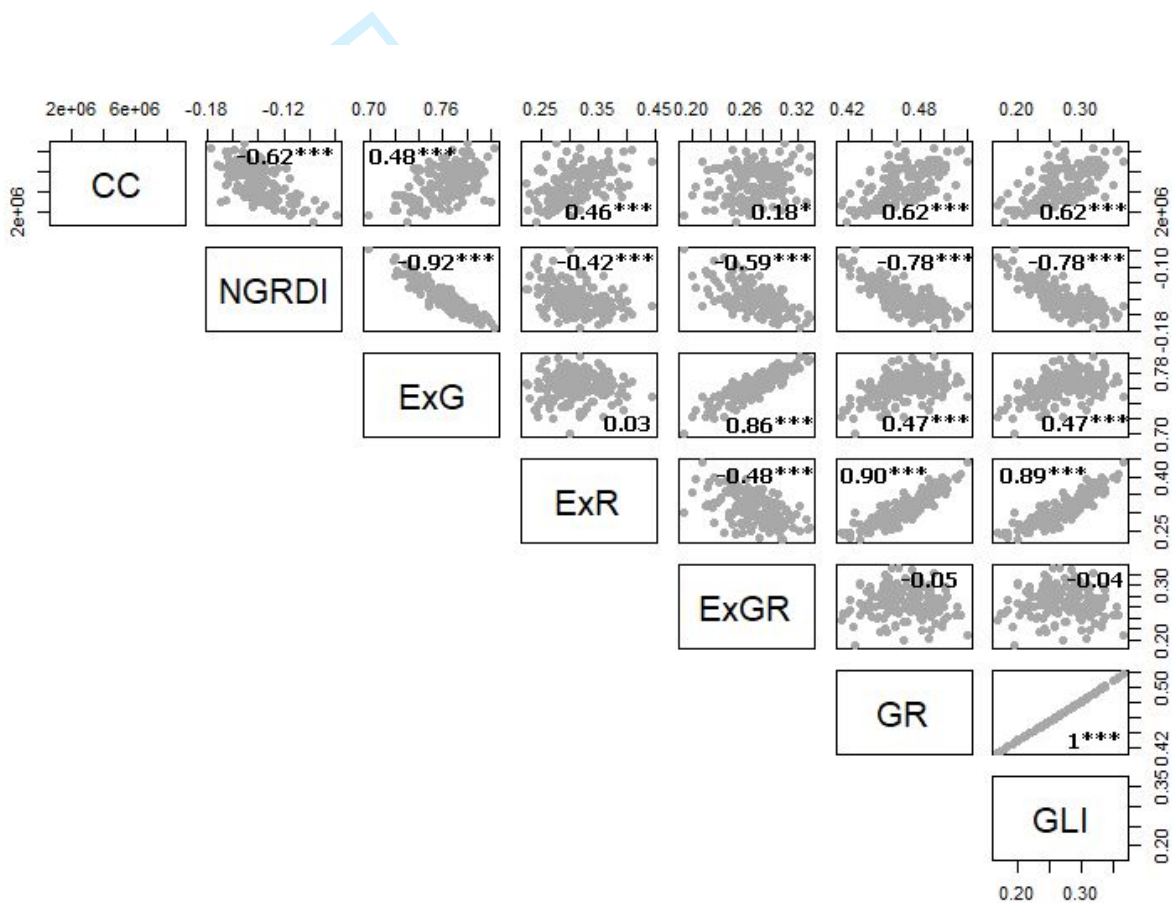
639

640

641

642 **Supplementary Fig 2.** Binary relationships and Pearson's correlation coefficients between seven  
 643 plant traits extracted from digital images of 200 *Urochloa* hybrids. CC= canopy cover, NGRDI=  
 644 normalized green red difference index, ExG= excess green index, ExR= excess red index, ExGR=  
 645 excess green minus excess red, GR= green ratio and GLI= green leaf index. Pearson's correlation  
 646 coefficients are indicated with their statistical significance as follows: \* $P \leq 0.1$ , \*\* $P \leq 0.01$ ,  
 647 \*\*\* $P \leq 0.001$ .

648



649

650

1 Proximal sensing of *Urochloa* grasses increases selection accuracy

2

3 Juan de la Cruz Jiménez<sup>1\*</sup>, Luisa Leiva<sup>2</sup>, Juan A. Cardoso<sup>3</sup>, Andrew N. French<sup>4</sup> and Kelly R. Thorp<sup>4</sup>

4

5 <sup>1</sup> UWA School of Agriculture and Environment, Faculty of Science, The University of Western Australia,  
6 35 Stirling Highway, Crawley, WA 6009, Australia.

7 <sup>2</sup> Department of plant breeding, Swedish University of Agricultural Sciences, Alnarp, Sweden.

8 <sup>3</sup> Tropical Forage Program, International Center for Tropical Agriculture (CIAT), Km 17 Recta Cali –  
9 Palmira, Colombia.

10 <sup>4</sup> USDA-ARS, U.S. Arid Land Agricultural Research Center, 21881 N Cardon Ln, Maricopa, AZ 85138,  
11 United States.

12

13

14

15 \* Corresponding author: Juan de la Cruz Jiménez: [juan.jimenezserna@research.uwa.edu.au](mailto:juan.jimenezserna@research.uwa.edu.au)

16

17

18

19

20

## 21 Abstract

22 In the American Tropics, livestock production is highly restricted by forage availability. In  
23 addition, the breeding and development of new forage varieties with outstanding yield and high  
24 nutritional quality is often limited by a lack of resources and poor technology. Non-destructive  
25 high throughput phenotyping offers a rapid and economical means to evaluate large numbers of  
26 genotypes. In this study, visual assessments, digital color images, and spectral reflectance data  
27 were collected from 200 *Urochloa* hybrids in a field setting. Partial least squares regression  
28 (PLSR) was applied to relate visual assessments, ~~vegetation indices~~digital image analysis and  
29 spectral data with shoot dry weight, ~~nitrogen (N) content~~ (DW), crude protein (CP) and chlorophyll  
30 content. Visual evaluations of biomass and greenness, digital color imaging, and hyperspectral  
31 canopy data were collected in 68, 40 and 80 minutes, respectively. Root mean squared errors of  
32 prediction for PLSR estimations of ~~dry weight~~, NDW, CP, and chlorophyll were lower for  
33 ~~vegetation indices~~ digital image analysis followed by hyperspectral analysis and visual  
34 assessments. This study showed that digital color image and spectral analysis techniques have the  
35 potential to improve precision and reduce time for tropical forage grass phenotyping.

36 Keywords: High throughput phenotyping, *Urochloa*, tropical forage grasses, plant breeding.

37

## 38 Introduction

39 Livestock productivity depends on forage availability and quality. Grasses from the *Urochloa* (syn.  
40 *Brachiaria*) genus have been widely planted in the tropics as forage for grazing ruminant livestock  
41 and are considered the most important forages in the American Tropics (Miles et al. 2004). The  
42 International Center for Tropical Agriculture (CIAT) in Colombia conducts a *Urochloa* breeding

43 program aimed at developing hybrids with outstanding performance on infertile, acidic soils with  
44 superior forage productivity and nutritional quality. The hybrid development process is difficult  
45 and time consuming. In a regular, three-year breeding cycle ~~(three years)~~, over 7000 hybrids are  
46 produced by open pollination, but fewer than 2% of these are retained for full evaluation.  
47 Approximately half of the population is discarded based on their reproductive mode (sexual  
48 ~~or~~ genotypes are discarded and apomictic hybrids are kept); another major proportion is discarded  
49 based on visual evaluations; and only a limited number of hybrids (approximately 100) are finally  
50 evaluated for different biotic and abiotic stresses (Valheria Castiblanco, personal communication).  
51 The evaluation of genotypes is restricted mainly by insufficient economic resources and lacking  
52 technology for rapid screening.

53 Periodic Forage grasses exhibiting great biomass production and high nutritional quality are key  
54 determinants of the productivity of grazing animals (Herrero et al., 2013). Therefore, evaluations  
55 of shoot biomass production and quality parameters (i.e. crude protein) are among the most  
56 important traits for improvement in any forage grasses breeding program. However, owing to the  
57 destructive nature of these measurements and the insufficient economic resources, the evaluation  
58 of these parameters is postponed to final stages of the breeding program characterized by a reduced  
59 number of genotypes. Instead of analytical measurements of forage quality and destructive  
60 biomass harvests, periodic visual evaluations of plant performance (i.e., plant biomass and  
61 greenness) over time ~~has been~~ traditionally used in ~~the CIAT's~~ *Urochloa* breeding  
62 ~~program~~ programs to select superior plants at initial stages of the breeding scheme (Miles et al.  
63 2004; Miles 2007). These visual evaluations are laborious and may not be sufficiently accurate  
64 especially in breeding populations characterized by high genetic diversity and substantial genotype  
65 x environment interaction (Walter et al. 2012). In this sense, the

66 The use of new technologies for in-field non-destructive, high throughput phenotyping (HTP),  
67 including digital image analysis and proximal hyperspectral sensing, ~~represents~~offers the  
68 possibility to precisely evaluate a larger number of genotypes than feasible in traditional ways,  
69 achieved at low cost, and implemented in a short period of time (Montes et al. 2007; White et al.  
70 2012; Andrade-Sanchez et al. 2014).

71 Proximal hyperspectral sensing provides continuous information along the visual and near-infrared  
72 electromagnetic spectrum. This information often relates to plant traits and has successfully been  
73 studied in grasses to estimate quality parameters (Skidmore et al. 2010; Pullanagari et al. 2012;  
74 Thulin et al. 2012; Ferner et al. 2015; Safari et al. 2016), diversity (Lopatin et al. 2017) and nutrient  
75 content (Fava et al. 2009; Knox et al. 2012; Ramoelo et al. 2013; Adjorlolo et al. 2015; Foster et  
76 al. 2017). Likewise, plant image analysis for phenotyping purposes is based on image  
77 segmentation to separate the soil background ~~(i.e., soil)~~ and the plant for further quantification of  
78 regions of interest (Tucker 1979; Woebbecke et al. 1995; Camargo 2004; Hunt et al. 2005). Digital  
79 image analysis has also been used for quantifying vegetation indices related to plant growth,  
80 greenness and nutritional status (Meyer and Camargo 2008; Hunt et al. 2013). Very few reports of  
81 hyperspectral (Numata et al. 2008) or image analysis of *Urochloa* grasses exist in literature  
82 (Jimenez et al. 2017).

83 No studies combining hyperspectral information and image analyses and comparing them to  
84 conventional phenotyping methods is available. Moreover, hyperspectral data have not been used  
85 to evaluate target traits in *Urochloa* breeding programs. In this study, in-field visual evaluations,  
86 proximal hyperspectral data, and digital imaging were collected over canopies of *Urochloa*  
87 hybrids. Partial least squares regression (~~PLSR~~) was used to relate hyperspectral information to  
88 field measurements and machine learning (i.e. naive Bayes multiclass) was used to extract



89 vegetation indices from overhead canopy images. The objectives of this study were to: 1) develop  
90 PLSR models for predicting NCP, forage ~~dry weight~~DW, and chlorophyll content; 2) ~~compute~~  
91 ~~vegetation indices~~extract plant traits from digital image analysis to relate with NCP, forage ~~dry~~  
92 ~~weight~~DW, and chlorophyll; and 3) demonstrate the superiority of HTP techniques as compared  
93 to conventional visual evaluation of traits. ~~Hyperspectral data or image analysis have not been~~  
94 ~~used to evaluate forage biomass, N or chlorophyll in Urochloa breeding programs. Moreover, no~~  
95 ~~studies combining hyperspectral information and image analysis and comparing it to conventional~~  
96 ~~phenotyping methods is available. Nitrogen, forage dry weight~~Crude protein, forage DW, and  
97 chlorophyll content were chosen as target traits in this study as they are key parameters  
98 determining both plant and cattle productivity. The development of HTP methodologies to  
99 evaluate tropical forages will increase the number of hybrids evaluated per selection cycle, thus  
100 permitting more intense selection and hence, genetic gain. The identification of new hybrids with  
101 outstanding performance (i.e. higher biomass, greener and high N-contentCP) will result in more  
102 productive pastures with concomitant increases in milk and meat production in livestock systems  
103 in tropical savannahs.

## 104 **Materials and methods**

### 105 *Field experiment*

106 Field data were obtained in August 2016 at the International Center for Tropical Agriculture  
107 (CIAT) in Cali, Colombia (Lat. 3° 29' N; Long. 76° 21' W; altitude 965 m). Four thousand  
108 *Urochloa* hybrids generated from crosses between the CIAT's *Urochloa* breeding program  
109 population SX12 and *U. decumbens* cv. Basilisk (CIAT 606) were initially planted in an andisol  
110 soil in an augmented block design and spaced at 1.5x1.5 m. These plants were visually evaluated  
111 four times (data not shown) for persistence, vigor and greenness after sequential cuttings every

112 three months for one year. After that period, 200 hybrids were randomly selected for further visual  
113 and HTP analysis. These 200 hybrids (~~and not, instead of~~ the entire population), were selected for  
114 economic and practical reasons. Visual evaluations of biomass and greenness, imaging and spectra  
115 collection were performed after 3 months ~~regrowth~~re-growth after cutting (see information below).  
116 Plant heights ranged from 20 to 50 cm and shoot architecture varied ~~from very prostrate to with~~  
117 both decumbent and erectus growth.

#### 118 *Visual evaluation*

119 Plant biomass was assessed using a nine-point visual scale, where level '9' indicated high shoot  
120 biomass with many tillers and leaves while level '1' indicated stunted growth with fewer tillers  
121 and leaves. Plant greenness was visually evaluated using a five-point visual scale, where level '5'  
122 represented intense dark green in all the leaves of the plant and level '1' indicated yellow-pale  
123 color in all leaves of the plant. This visual evaluation was conducted in 68 minutes one week before  
124 the HTP measurements (Table 1).

#### 125 *Imaging collection and analysis*

126 Individual, digital color images for each of the 200 hybrids were taken at 1.2 m above the soil  
127 surface using a commercial digital 13-Megapixel camera (Coolpix P6000, Nikon, Japan) fixed to  
128 a buggy tractor. Digital images were saved in 4224 x 3168 pixel JPG format ~~and vegetation indices~~  
129 ~~were analyzed.~~ The canopy cover (CC) and six vegetation indices including the normalized green  
130 red difference index (NGRDI), excess green index (ExG), excess red index (ExR), excess green  
131 minus excess red (ExGR), green ratio (GR) and green leaf index (GLI) were created using the  
132 formulae as indicated in Table 2. The canopy cover was extracted by dividing the total number of  
133 pixels representing the plant by the total number of pixels in each image. The vegetation indices  
134 were extracted using naive Bayes multiclass. Briefly, the distribution of colors in a set of digital  
135 color images (training set) was used to estimate the probability density function for each of the



136 different region of interest (i.e. plant and background). Once the regions of interest were defined,  
 137 in the training set, the machine learning process was applied to all images to accurately classify  
 138 and separate regions of interest; ~~therefore.~~ Therefore, every ~~new~~-pixel in an image was classified  
 139 into the previously defined plant and background classes. Every pixel characterizing the plant (but  
 140 not the background) was then decomposed into red, (R), green, (G), and blue (RGBB) channels  
 141 ~~for.~~ These channels were then normalized as follows:

$$143 \quad r \equiv \frac{R}{R + G + B}; g \equiv \frac{G}{R + G + B}; b \equiv \frac{B}{R + G + B}$$

144  
 145 Normalization makes the variations of light intensities uniform across the spectral distribution,  
 146 thus, the individual color components (i.e. r,g,b) are independent from the overall brightness of  
 147 the image (Cheng et al. 2011). Normalized channels were further used for the quantification of the  
 148 vegetation indices. ~~Seven vegetation indices including canopy cover, normalized red green~~  
 149 ~~difference index (Tucker 1979), excess green index (Woebbecke et al. 1995), excess red index~~  
 150 ~~(Meyer et al. 1998), excess green minus excess red (Camargo 2004), green ratio, and green leaf~~  
 151 ~~index (Louhaichi et al. 2001) were calculated. (Table 2).~~ Image analysis code was written in Java  
 152 and run in ImageJ software (National Institutes of Health, Bethesda, Maryland, USA). Images  
 153 were ~~collecting~~collected early in the morning to avoid beam solar radiation interferences. Digital  
 154 images contained the whole plant in addition to the 23-cm diameter field-of-view (as indicated  
 155 below for hyperspectral measurements, Supplementary Fig 1). The collection process took 40  
 156 minutes (Table 1).

157  
 158 *Spectral collection and analysis*

159 Hyperspectral field data collections were performed on clear days at full sun exposure around 11  
160 am by positioning a hand-held field spectroradiometer (Fieldspec 2, Malvern Panalytical, Malvern,  
161 UK) directly above the plant canopy. The instrument was used with no foreoptics, which provided  
162 a 25-degree full conical angle field-of-view. To avoid soil background noise, the bare optical input  
163 was positioned at 50 cm from the top of the plant canopy to yield a 23-cm diameter field of view.  
164 The instrument collected information in 750 narrow wavebands from 325 to 1075 nm in 1 nm  
165 intervals. One or ten spectral scans were collected per plant and 50 plants were evaluated daily in  
166 about 20 minutes. Differences in the collection protocols were ~~deliberately done for comparison~~  
167 ~~purposes tested to evaluate the most effective way.~~ Different spectra collection  
168 ~~paradigms processes~~ (1 or 10 scans) did not yield significant differences in the root mean squared  
169 error of prediction for the different traits evaluated (Supplementary ~~FigTable~~ 1). Radiometric  
170 collections over a 99% Spectralon panel (Labsphere, Inc., North Sutton, New Hampshire) were  
171 used to describe incoming solar irradiance throughout the data collection process. The radiometric  
172 collections over the calibration panel were made before starting and after every five canopy scans  
173 or when slight changes in solar irradiance ~~due to cloud cover~~ occurred. The values of the  
174 Spectralon panel radiance were used to compute the canopy reflectance of the plants in each  
175 wavelength over the time of spectra collection. Subsequently, 401 bands from 500 to 900 nm were  
176 used for analysis. Based on visual inspection of reflectance spectra, these bands were typically less  
177 noisy, as compared to bands at the bounds of detector sensitivity. Spectral collection process was  
178 run in 80 minutes (Table 1).

### 179 *Laboratory sample collections*

180 Plants were immediately harvested after spectra collection. Aboveground tissue was removed by  
181 cutting the area defined by a 23-cm diameter plastic circle co-located with the spectral data

182 collection area. Tissues were packed in plastic bags and stored on ice in a cooler in the field and  
183 then transported to the laboratory. The extraction of chlorophyll was performed by adding 100 mg  
184 of fresh tissue to 80% (v/v) cold methanol, and the mix was homogenized using a pestle in a mortar  
185 until the plant residue was clear and the solution was uniform. This solution was then filtered and  
186 absorbance was determined with a spectrophotometer (Synergy HT, Biotek, Winooski, USA).  
187 Total chlorophyll concentration was calculated according to Lichtenthaler and Welburn (1983).  
188 Dry weight (DW) was measured on an electronic balance (PB602S, Mettler Toledo, LLC,  
189 Columbus, OH, USA) after oven-drying the samples for three days at 60 °C. Nitrogen  
190 concentrations in the dry tissue were determined by using an automated nitrogen-carbon analyser  
191 (Sercon, Crewe, UK). *Urochloa* and common bean (*Phaseolus vulgaris*) leaves were used as  
192 reference tissues for confirmation of the reliability of the analyses. The crude protein content was  
193 calculated by multiplying nitrogen content with 6.25, as protein is assumed to contain 16%  
194 nitrogen on average.

195

#### 196 *Statistical analysis*

197 Visual evaluations, ~~vegetation indices~~ digital image analysis, spectral reflectance, and plant trait  
198 data were incorporated into a partial least squares regression (PLSR) algorithm (Mevik and  
199 Wehrens 2007) within the R Project for Statistical Computing (<http://www.r-project.org>). Models  
200 were developed to estimate/predict each plant trait (i.e. CP, DW and chlorophyll) and to compare  
201 the precision for prediction of each of the different methods of phenotyping. Partial least squares  
202 regression was used in preference to conventional least squares analysis to reduce co-linearity  
203 effects. Thorp et al. (2011) provided the details on the PLSR methodology used in the present  
204 study. Briefly, if  $Y$  is an  $n \times 1$  vector of responses (i.e. N, dry weight CP, DW or chlorophyll content)

205 and  $\mathbf{X}$  is an  $n$ -observation by  $p$ -variable matrix of predictors (a set of visual evaluations, vegetation  
206 indicesdigital image analysis, or spectral reflectance data), PLSR aims to decompose  $\mathbf{X}$  into a set  
207 of  $A$  orthogonal scores such that the covariance with corresponding  $\mathbf{Y}$  scores is maximized. The  
208  $X$ -weight and  $Y$ -loading vectors that result from the decomposition are used to estimate the vector  
209 of regression coefficients,  $\beta_{\text{PLS}}$ , such that

$$210 \mathbf{Y} = \mathbf{X} \beta_{\text{PLS}} + \boldsymbol{\varepsilon}$$

211 where  $\boldsymbol{\varepsilon}$  is an  $n \times 1$  vector of error terms.

212 Leave-one-out cross validation was used to test model predictions for independent data. Results  
213 were reported for PLSR models with the number of factors that minimized the root mean squared  
214 error of cross validation. Pearson's correlation coefficients were calculated for the different traits  
215 extracted from digital color images taken from *Urochloa* hybrids.

## 216 Results

217 In this study, visual evaluations of biomass and greenness, digital color imaging and hyperspectral  
218 data were collected on 200 *Urochloa* hybrids in 68, 40 or 80 minutes, respectively (Table 1). High  
219 variability for the different characteristics of dry weight, nitrogenDW, CP and chlorophyll content  
220 evaluated on 200 *Urochloa* hybrids was found (Table 23).

### 221 *Visual assessments*

222 Partial least squares regressions for measured traits of DW, NCP and chlorophyll and based on  
223 visual evaluations of biomass and greenness performed with a root mean square error of prediction  
224 (RMSEP) of 8.47 g plant<sup>-1</sup>, 1.76% and 0.60 mg g FW respectively (Fig 1).

### 225 *Spectral data and digital image phenotyping*

226 The PLSR models developed from ~~seven vegetation indices~~ the digital image analysis estimated  
227 DW, NCP and chlorophyll with a RMSEP of 7.7981 g plant<sup>-1</sup>, 1.53% and 0.57 mg g FW,  
228 respectively (Fig 2). Differences on the correlation coefficients among traits extracted from image  
229 analysis indicated that including different indices into the model added independent information  
230 to build stronger PLSR models (Supplementary Fig 2). The contribution of each trait extracted  
231 from digital image analysis to the overall prediction of each destructively-measured trait is shown  
232 in Table 4. The GLI had the stronger positive influence on the PLSR model for predicting DW.  
233 The ExGR had the stronger positive influence on the PLSR model for predicting both CP and  
234 chlorophyll content.

235 The fitted PLSR models developed from 401 wavebands of canopy spectral reflectance estimated  
236 DW, NCP and chlorophyll with a RMSEP of 7.90 g plant<sup>-1</sup>, 1.63% and 0.55 mg g FW, respectively  
237 (Fig 3).

238 The contribution of each spectral waveband to the overall prediction of each destructively-  
239 measured trait is shown in the Fig- 4. In the PLSR model for DW, ~~three bands characterized the~~  
240 ~~dry weight of Urochloa. Local~~ local extrema in regression coefficients were found at 543, 668701  
241 and 744674 nm, corresponding to ~~visible green light,~~ red light near the inflection band and NIR  
242 radiation ~~red light~~, respectively (Fig- 4a). Strong positive contribution to ~~dry weight~~ DW estimation  
243 were with ~~green light (543) and~~ NIR (744700-750), and a strong negative contribution with red  
244 light (668674-640). In the PLSR models for NCP and chlorophyll, regression coefficient plots  
245 exhibited ~~a noisy pattern with less defined extrema.~~ strong positive contribution for traits estimation  
246 in the visible green light (Fig 4b and c). The PLSR models for NCP contrasted wavebands in the  
247 visible spectrum with positive contribution from wavebands around 513503 nm and negative  
248 contributions from wavebands at 676678 nm. ~~Wavebands at 600 nm and in the NIR contributed~~

249 ~~less to the model for N (Fig. 4b). Regression~~ Similarly, regression coefficients for total chlorophyll  
250 indicated strong positive contribution ~~from NIR wavelengths and strong in the visible spectrum~~  
251 ~~around 504 nm and~~ negative contribution throughout the visible wavebands, especially ~~from 525~~  
252 ~~to 625 nm, and at the red edge at 705/643 nm~~ (Fig. 4c). This is sensible considering visible light  
253 absorption is increased with additional leaf chlorophyll. ~~In the PLSR model for chlorophyll, local~~  
254 ~~extrema in regression coefficients were found at 567, 674, 705 and 763 nm, which correspond to~~  
255 ~~green light at the edge of yellow, red light, red light near the red inflection band and NIR radiation.~~

## 256 Discussion

257 The results from this study demonstrate that the current visual assessment methodology at initial  
258 steps of the breeding cycle in the CIAT *Urochloa* breeding program can be improved ~~by the use~~  
259 ~~of using~~ non-destructive ~~high-throughput phenotyping~~ HTP techniques. ~~The use of color~~ Color  
260 imaging, hyperspectral analysis, and PLSR models is more precise and faster than visual  
261 evaluations, thus increasing the number of plants evaluated in the tropical forage breeding  
262 program.

263 Visual evaluations of plant growth and greenness (characteristics associated with N content, and  
264 therefore CP and chlorophyll concentration in leaves) have traditionally been used to discard  
265 *Urochloa* hybrids at initial stages of plant phenotyping. The visual evaluation of an entire breeding  
266 population (i.e., ~~4000~~ 7,000 hybrids) is a slow, costly and tedious process, and is often biased by  
267 subjectivity and human fatigue, especially when phenotypic variation of such traits is high (Table  
268 23). In this study, the estimation of DW, NCP, and chlorophyll content was more precisely and  
269 consistently estimated by HTP techniques. Dry weight and NCP predictions were more accurate  
270 using ~~vegetation indices~~ digital image analysis, followed by spectral analysis and visual  
271 evaluations. Chlorophyll content was better estimated by the analysis of 401 spectral wavebands,

272 followed by color image analysis and finally visual evaluations (Fig 1, 2 and 3). ~~Likewise, the~~The  
273 time required to run non-destructive HTP evaluations was considerably shorter by 28 minutes per  
274 200 plants for color image analysis than visual evaluations, but longer by 12 minutes per 200 plants  
275 in hyperspectral than in visual evaluations (Table 1).

276 The moderate trends in the relationship between *Urochloa* canopy imaging and reflectance and  
277 measured DW, ~~NCP~~ and chlorophyll may indicate that the method is not appropriate for very  
278 precise estimations of these traits. However, for breeding purposes where a large percentage of  
279 hybrids are discarded without detailed evaluation due to scarce resources, a difference in DW of  
280 7.90 g plant<sup>-1</sup> or a difference of 1.63% in the ~~NCP~~ content of plants may be acceptable during  
281 initial stages of plant breeding. This moderate trend between *Urochloa* canopy analysis and  
282 measured traits in this study can be explained by dissimilarities in the ~~*Urochloa* genotypes~~-canopy  
283 architecture of the *Urochloa* genotypes (Numata et al. 2008), as well as different growth patterns  
284 during recovery from cutting. The further evaluation of breeding populations with contrasting  
285 canopy architecture will improve the accuracy of the PLSR model to predict the targeted traits.  
286 Nonetheless, by combining both digital image and hyperspectral analysis techniques, higher  
287 precision accuracy for DW, CP and Chlorophyll content can be achieved.

288 The vegetation indices (see ~~materials and methods~~Table 2) extracted from color images of 200  
289 *Urochloa* hybrids were originally developed to ~~extract~~separate green plants from the background  
290 by extracting green and red colors from ~~the image data to estimate~~digital images. These indices  
291 have been related to different plant characteristics including biomass, chlorophyll content and ~~the~~  
292 nutritional status ~~of plants~~ (Tucker 1979; Woebbecke et al. 1995; Camargo 2004; Hunt et al. 2005;  
293 Meyer and Camargo 2008; Hunt et al. 2013; Lee and Lee 2013; Wang et al. 2013). -In this study,  
294 ~~vegetation indices~~ digital image analysis performed better than hyperspectral scanning analysis to

295 estimate DW and NCP (Fig 2 and 3). Nonetheless, the use of spectral analysis over grasses  
296 becomes more important when this technique is used to detect either nutritional or anti-nutritional  
297 compounds (i.e. proteinmetabolisable energy, digestibility, fiber) that are better estimated with the  
298 near-infrared regions of the electromagnetic spectra (Curran 1989; Pullanagari et al. 2012; Ferner  
299 et al. 2015). In this sense, the use of digital color image analysis and hyperspectral analysis is  
300 complementary because by using both techniques a diverse set of plant traits can accurately be  
301 predicted- and by adding extra factors to the prediction model, higher prediction accuracy can be  
302 achieved (cf. Numata et al. 2008). Future efforts will use data mining to fine-tune the spectral  
303 bands included in the PLSR model (Thorp et al. 2017), which can reduce model error and improve  
304 model fit statistics. Although testing multiple methods of analysis was not the intention of this  
305 study, future research could also test other techniques (e.g., artificial neural networks) for relating  
306 HTP measurements to plant traits.

307 The regression coefficients for the PLSR for DW and chlorophyll content obtained in this study  
308 highlight that the key wavelengths for the prediction of these traits ~~were located~~occur in the green,  
309 red, red-edge and NIR regions of the electromagnetic spectrum (Fig 4). Previous hyperspectral  
310 studies have highlighted those regions as being highly representative for dry mass and chlorophyll  
311 content in plants (Lichtenthaler et al. 1996; Thenkabail et al. 2000; Mutanga and Skidmore 2004;  
312 Fava et al. 2009; Thorp et al. 2011; Adjorlolo et al. 2015; Dou et al, 2018). Although some  
313 similarities were found between wavebands among the different traits, the general regression  
314 coefficients differed among the traits, thus demonstrating that the reflectance data in a given  
315 waveband contributed differently toward the estimation of a given trait. Given the logistical burden  
316 to collect and analyze hyperspectral scans, the identification of informative key bands associated  
317 with each evaluated trait can improve the HTP process (Thorp et al. 2017). Results from this study



318 will help guide selection of optimal bands in the construction of multispectral sensors tailored to  
319 predict specific traits of interest in tropical forage breeding programs.

320 The PLSR models for predicting DW, CP and chlorophyll content can be now used to evaluate the  
321 next generation of hybrids from the same *Urochloa* gene pool (i.e. *U. ruziziensis* – *U. brizantha* –  
322 *U. decumbens*). The accuracy of this prediction models relies on collection protocols similar to the  
323 explained in the Materials and Methods section and evaluations on plants with comparable growth  
324 characteristics as the hybrids evaluated here (i.e. about three months after regrowth). The  
325 prediction accuracy will likely be reduced on larger plants with higher biomass (Hill 2004) and a  
326 greater proportion of senescent leaves (Asner 1998). The development of more precise PLSR  
327 models to predict variables of interest in a breeding program requires an ongoing effort. The  
328 collection of ground data every year while making improvements to standardize collection  
329 protocols and incorporate wider range of genotypes will result in more accurate and robust models.  
330 Larger data sets will increase estimation precision.

331

## 332 **Conclusions**

333 In this study, 200 *Urochloa* hybrids were ~~successfully~~ monitored in 40 and 80 minutes by digital  
334 imaging and spectral analysis, respectively (Table 1). At this pace, more than 1000 *Urochloa*  
335 hybrids ~~can~~ be evaluated in a period of less than 7 hours. This means ~~more~~ that forage biomass  
336 and quality in a high number of genotypes ~~could~~ would be reliably evaluated with minimal  
337 increased acquisition costs (~~compared~~ relative to destructive harvest). This demonstrates the  
338 superiority of HTP techniques as compared to conventional visual evaluation of traits. The PLSR  
339 models for predicting CP, forage DW, and chlorophyll content developed in this study supports

340 ~~the evaluation of higher numbers of genotypes at initial stages of the breeding program.~~ The greater  
341 ~~numbernumbers~~ of plants evaluated reliably every year in the *Urochloa* breeding program, the  
342 greater the genetic gain will be. Therefore, the use of image analysis and hyperspectral monitoring  
343 over *Urochloa* hybrids canopies will benefit the on-going breeding program. ~~Likewise, the~~The  
344 application of this ~~methodology~~HTP method could be of great help in rural remote areas ~~without~~  
345 ~~appropriate~~lacking facilities to perform destructive harvest and plant chemical analysis. ~~Additional~~  
346 ~~studies on Urochloa plants with contrasting architectures need to be performed to optimize PLSR~~  
347 ~~models. Moreover, more careful field measurements over plants with similar regrowth capacity~~  
348 ~~are required~~Research is underway to improve the ~~prediction models.~~ Furthermore, utility of  
349 proximal sensing by considering a greater range of canopy architectural configurations and  
350 evaluating the potential to assess nutritional quality ~~traits,~~ including ~~protein~~characteristics such as  
351 metabolisable energy, fiber, digestibility ~~and non-digestible fractions of the forage (~~ lignin and  
352 cellulose) ~~must be evaluated through proximal hyperspectral sensing to improve phenotyping~~  
353 fractions in *Urochloa* grasses.

354

### 355 **Conflict of interest**

356 The authors have no conflicts of interest to declare.

### 357 **Acknowledgements**

358 We thank Dr. John W. Miles for helpful suggestions and comments to early version of this  
359 manuscript. JCJ is grateful to the USDA Foreign Agricultural Service, Norman Borlaug  
360 Fellowship Program for a training fellowship. JCJ thanks the CIAT's Young Scientist Award 2016  
361 Program for travel assistance. This work was partially undertaken as part of the CGIAR Research

362 Program on Livestock. We thank all donors that globally support our work through their  
363 contributions to the CGIAR system.

## 364 **References**

365 Adjorlolo C, Mutanga O, Cho MA (2015) Predicting C3 and C4 grass nutrient variability using *in*  
366 *situ* canopy reflectance and partial least squares regression. *International Journal of Remote*  
367 *Sensing* **36(6)**, 1743–1761.

368

369 Andrade-Sanchez P, Gore MA, Heun JT, Thorp KR, Carmo-Silva AE, French AN, Salvucci ME,  
370 White JW (2014) Development and evaluation of a field-based high throughput phenotyping  
371 platform. *Functional Plant Biology* **41**, 68–79.

372

373 Camargo JN (2004) A combined statistical—soft computing approach for classification and  
374 mapping weed species in minimum tillage systems. University of Nebraska, Lincoln, NE

375

376 Currant PJ (1989) Remote sensing of foliar chemistry. *Remote Sensing of the Environment* **30**,  
377 271–278.

378

379 Dou Z, Cui L, Li J, Zhu Y, Gao C, Pan X, Lei Y, Zhang M, Zhao X, Li W (2018) Hyperspectral  
380 estimation of the chlorophyll content in short-term and long-term restorations of mangrove in  
381 Quanzhou Bay Estuary, China. *Sustainability* **10**, 1127; doi:10.3390/su10041127

382

383 Fava F, Colombo R, Bocchi S, Meroni M, Sitzia M, Fois N, Zucca C (2009) Identification of  
384 hyperspectral vegetation indices for Mediterranean pasture characterization. *International Journal*  
385 *of Applied Earth Observation and Geoinformation* **11**, 233–243.

386  
387 Ferner J, Linstadter A, Sudekum KH, Schmidlein S (2015) Spectral indicators of forage quality  
388 in West Africa's tropical savannas. *International Journal of Applied Earth Observation and*  
389 *Geoinformation* **41**, 99–106.

390  
391 Foster AJ, Kakani VG, Mosali J (2017) Estimation of bioenergy crop yield and N status by  
392 hyperspectral canopy reflectance and partial least square regression. *Precision Agriculture* **18**,  
393 192–209.

394  
395 Herrero M, Havlik P, Valin H, Notenbaert A, Rufino MC, Thornton PK, Blümmel M, Weiss F,  
396 Grace D, Obersteiner M (2013) Biomass use, production, feed efficiencies, and greenhouse gas  
397 emissions from global livestock systems. *Proceedings of the National Academy of Science of the*  
398 *United States of America* **110**(52), 20888–20893.

399  
400 Hunt Jr ER, Cavigelli M, Daughtry CST, McMurtrey J, Walthall CL (2005) Evaluation of digital  
401 photography from model aircraft for remote sensing of crop biomass and nitrogen status. *Precision*  
402 *Agriculture* **6**, 359–378.

403  
404 Hunt Jr ER, Doraiswamy PC, McMurtrey J, Daughtry CST, Perry EM, Akhmedov B (2013) A  
405 visible band index for remote sensing leaf chlorophyll content at the canopy scale. *International*  
406 *Journal of Applied Earth Observation and Geoinformation* **21**, 103–112.

407

408 Jiménez JC, Cardoso JA, Leiva LF, Gil J, Forero MG, Worthington ML, Miles JW, Rao IM (2017)

409 Non-destructive phenotyping to identify brachiaria hybrids tolerant to waterlogging stress under

410 field conditions. *Frontiers in Plant Science* **8**, 167. doi: 10.3389/fpls.2017.00167

411

412 Knox NM, Skidmore AK, Prins HHT, Heitkönig IMA, Slotow R, van der Waal C, de Boer WF

413 (2012) Remote sensing of forage nutrients: Combining ecological and spectral absorption feature

414 data. *ISPRS Journal of Photogrammetry and Remote Sensing* **72**, 27–35.

415

416 Meyer GE, Hindman TW, Lakshmi K (1998) Machine vision detection parameters for plant

417 species identification. In ‘Precision Agriculture and Biological Quality’. (Eds GE Meyer, JA De

418 Shazer) pp. 327–335. (Proceedings of SPIE. vol. 3543, Bellingham, WA)

419

420 Meyer GE, Camargo J (2008) Verification of color vegetation indices for automated crop imaging

421 applications. *Computers and Electronics in Agriculture* **63**, 282–293.

422

423 Mevik BH, Wehrens R (2007) The pls Package: Principal component and partial least squares

424 regression in R. *Journal of Statistical Software* **18(2)**, 1-23.

425

426 Miles JW, Do Valle CB, Rao IM, Euclides VPB (2004) Brachiariagrasses. American Society of

427 Agronomy, Crop Science Society of America, Soil Science Society of America, 677 S. Segoe Rd.,

428 Madison, WI 53711, USA. *Warm Season (C4) Grasses*, Agronomy Monograph no. 45.

429

430 Miles JW (2007) Apomixis for cultivar development in tropical forage grasses. *Crop Science*  
431 **47(S3)**, S238–S249.

432

433 Montes JM, Melchinger AE, Reif JC (2007) Novel throughput phenotyping platforms in plant  
434 genetic studies. *Trends in Plant Science* **12**, 433–436.

435

436 Mutanga O, Skidmore AK (2004) Narrow band vegetation indices overcome the saturation  
437 problem in biomass estimation. *International Journal of Remote Sensing* **25(19)**, 3999–4014.

438

439 Numata I, Roberts DA, Chadwick OA, Schimel JP, Galvão LS, Soares JV (2008) Evaluation of  
440 hyperspectral data for pasture estimate in the Brazilian Amazon using field and imaging  
441 spectrometers. *Remote Sensing of Environment* **112**, 1569–1583.

442

443 Lee KJ, Lee BW (2013) Estimation of rice growth and nitrogen nutrition status using color digital  
444 camera image analysis. *European Journal of Agronomy* **48**, 57–65.

445

446 Lichtenthaler HK, Wellburn AA (1983) Determination of total carotenoids and chlorophylls a and  
447 b of leaf extracts in different solvents. *Biochemical society transactions* **603**, 591–592.

448

449 Lichtenthaler HK, Gitelson A, Lang M (1996) Non-destructive determination of chlorophyll  
450 content of leaves of a green and an aurea mutant of tobacco by reflectance measurements. *Journal*  
451 *of Plant Physiology* **148**, 483–493.

452

- 453 Lopatin J, Fassnacht FE, Kattenborn T, Schmidtlein S (2017) Mapping plant species in mixed  
454 grassland communities using close range imaging spectroscopy. *Remote Sensing of Environment*  
455 **201**, 12–23.
- 456
- 457 Louhaichi M, Borman MM, Johnson DE (2001) Spatially located platform and aerial photography  
458 for documentation of grazing impacts on wheat. *Geocarto International* **16(1)**, 65-70.
- 459
- 460 Pullanagari RR, Yule IJ, Tuohy MP, Hedley MJ, Dynes RA, King WM (2012) In-field  
461 hyperspectral proximal sensing for estimating quality parameters of mixed pasture. *Precision in*  
462 *Agriculture* **13**, 351–369.
- 463
- 464 Ramoelo A, Skidmore AK, Schlerf M, Heitkönig IWA, Mathieu R, Cho MS (2013) Savanna grass  
465 nitrogen to phosphorous ratio estimation using field spectroscopy and the potential for estimation  
466 with imaging spectroscopy. *International Journal of Applied Earth Observation and*  
467 *Geoinformation* **23**, 334–343.
- 468
- 469 Safari H, Fricke T, Wachendorf M (2016) Determination of fibre and protein content in  
470 heterogeneous pastures using field spectroscopy and ultrasonic sward height measurements.  
471 *Computers and Electronics in Agriculture* **123**, 256–263.
- 472
- 473 Skidmore AK, Ferwerda JG, Mutanga O, Van Wieren SE, Peel M, Grant RC, Prins HHT, Balciik  
474 FB, Venus V (2010) Forage quality of savannas - Simultaneously mapping foliar protein and  
475 polyphenols for trees and grass using hyperspectral imagery. *Remote Sensing of Environment* **114**,  
476 64–72.

- 477
- 478 Thenkabail PS, Smith RB, Pauw ED (2000) Hyperspectral vegetation indices and their  
479 relationships with agricultural crop characteristics. *Remote Sensing of Environment* **71**, 158–182.
- 480
- 481 Thorp KR, Dierig DA, French AN, Hunsaker DJ (2011) Analysis of hyperspectral reflectance data  
482 for monitoring growth and development of lesquerella. *Industrial Crops and Products* **33**, 524–  
483 531.
- 484
- 485 Thorp KR, Wang G, Bronson KF, Badaruddin M, Mon J (2017) Hyperspectral data mining to  
486 identify relevant canopy spectral features for estimating durum wheat growth, nitrogen status, and  
487 grain yield. *Computers and Electronics in Agriculture* **136**, 1-12.
- 488
- 489 Thulin S, Hill MJ, Held A, Jones S, Woodgate P (2012) Hyperspectral determination of feed  
490 quality constituents in temperate pastures: Effect of processing methods on predictive relationships  
491 from partial least squares regression. *International Journal of Applied Earth Observation and*  
492 *Geoinformation* **19**, 322–334.
- 493
- 494 Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation.  
495 *Remote sensing of environment* **8**, 127-150.
- 496
- 497 Walter A, Studer B, Kölliker R (2012) Advanced phenotyping offers opportunities for improved  
498 breeding of forage and turf species. *Annals of Botany* **110**, 1271–1279.
- 499



500 Wang Y, Wang D, Zhang G, Wang J (2013) Estimating nitrogen status of rice using the image  
501 segmentation of G-R thresholding method. *Field Crops Research* **149**, 33–39.

502

503 Woebbecke DM, Meyer GE, Von Bargen K, Mortensen DA (1995) Color indices for weed  
504 identification under various soil, residue, and lighting conditions. *Transactions of the ASAE* **38(1)**,  
505 259-269.

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

For Review Only

524

525

526 **Table 1.** Phenotyping techniques used in the present study, the time of evaluation, its application,  
 527 advantages and disadvantages.

528

Phenotyping technique	Time of evaluation*	Applications	Advantages	Disadvantages
Visual evaluation	68 min	Visual observations of different plant characteristics	Easy operation, low cost, evaluations can be performed under diverse conditions and environments	Evaluation of low number of genotypes, evaluation subjected to human bias and fatigue
Image analysis	40 min	Quantification of canopy cover and vegetation indices in the visible electromagnetic spectrum	Easy operation, low cost, greater number of plants evaluated, determination of several vegetation and water indices	Changes in ambient light conditions limit calculation of vegetation indices, data analysis is moderately complex
Hyperspectral analysis	80 min	Canopy reflectance information in the visible and near infra-red regions of the electromagnetic spectrum. Information can be used to predict biochemical <del>compositions</del> <u>composition</u> of plants	Moderately easy operation, greater number of plants evaluated, determination of nutritional and biochemical composition of leaf/canopy	Low solar radiation or cloudy days limit analysis, sensor and white reference calibration is frequently needed, data analysis is complex

529 \* The time of evaluation refers to 200 *Urochloa* plants evaluated under the conditions of the  
 530 present study.

531

532

533

534

535

536

537

538

539 **Table 2.** Canopy cover and vegetation indices calculated from digital images of 200 *Urochloa*  
 540 hybrids. Vegetation indices were extracted using a naive Bayes multiclass machine learning  
 541 approach. Indices were then incorporated into a PLSR model to predict crude protein, dry weight  
 542 biomass and chlorophyll content.

543

<u>Plant traits</u>	<u>Name</u>	<u>Formula*</u>	<u>Reference</u>
<u>CC**</u>	<u>Canopy cover</u>	<u>Nc/Nt</u>	<u>=</u>
<u>NGRDI</u>	<u>Normalized green red difference index</u>	<u>(g-r)/(g+r)</u>	<u>Hunt et al., 2005</u>
<u>ExG</u>	<u>Excess green index</u>	<u>2g-r-b</u>	<u>Woebbecke et al., 1995</u>
<u>ExR</u>	<u>Excess red index</u>	<u>1.3r-g</u>	<u>Meyer et al., 1998</u>
<u>ExGR</u>	<u>Excess green minus excess red</u>	<u>ExG-ExR</u>	<u>Camargo 2004</u>
<u>GR</u>	<u>Green ratio</u>	<u>g/(r+g+b)</u>	<u>Tucker 1979</u>
<u>GLI</u>	<u>Green leaf index</u>	<u>(2g-r-b)/(2g+r+b)</u>	<u>Louhaichi et al. 2001</u>

\*r, g and b denote the normalized pixel values of each channel on the RGB colour mode.

\*\* No normalization was performed for the canopy cover quantification. Nc= total number of pixels representing the canopy, Nt= total number of pixels in the picture.

544

545

546

547

548 **Table 3.** Plant traits measured in 200 *Urochloa* hybrids.

549

Trait	Min	Max	Mean	CV (%)
Dry Weight (g plant <sup>-1</sup> )	6.74	64.1	30.22	34.81
<del>Nitrogen</del> Crude Protein (%)	6.76	21.58	11.23	19.68
Chlorophyll (mg g FW)	0.87	6.41	2.88	24.31

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565 **Table 4.** Regression coefficients of the fitted partial least square regression models of seven traits  
 566 extracted from digital image analysis. Positive and negative coefficients indicate positive and  
 567 negative influence on the prediction model, respectively.

<u>Traits*</u>	<u>Dry weight (g.plant<sup>-1</sup>)</u>	<u>Crude protein (% DW)</u>	<u>Chlorophyll (mg.g<sup>-1</sup>)</u>
<u>CC</u>	<u>3.760545</u>	<u>-0.23114345</u>	<u>-0.01673706</u>
<u>NGRDI</u>	<u>9.948634</u>	<u>0.08600517</u>	<u>-0.03532585</u>
<u>ExG</u>	<u>-14.3163</u>	<u>0.07760486</u>	<u>0.0730642</u>
<u>ExR</u>	<u>-32.126212</u>	<u>-0.39106455</u>	<u>-0.07758547</u>
<u>ExGR</u>	<u>3.724492</u>	<u>0.26592971</u>	<u>0.10326555</u>
<u>GR</u>	<u>-34.770799</u>	<u>-0.31158671</u>	<u>-0.03624834</u>
<u>GLI</u>	<u>80.87152</u>	<u>-0.31108316</u>	<u>-0.0357783</u>

568 \* CC= canopy cover, NGRDI= normalized green red difference index, ExG= excess green index,  
 569 ExR= excess red index, ExGR= excess green minus excess red, GR= green ratio and GLI= green  
 570 leaf index.  
 571

572

573

574

575

576

577

578

579

580

581

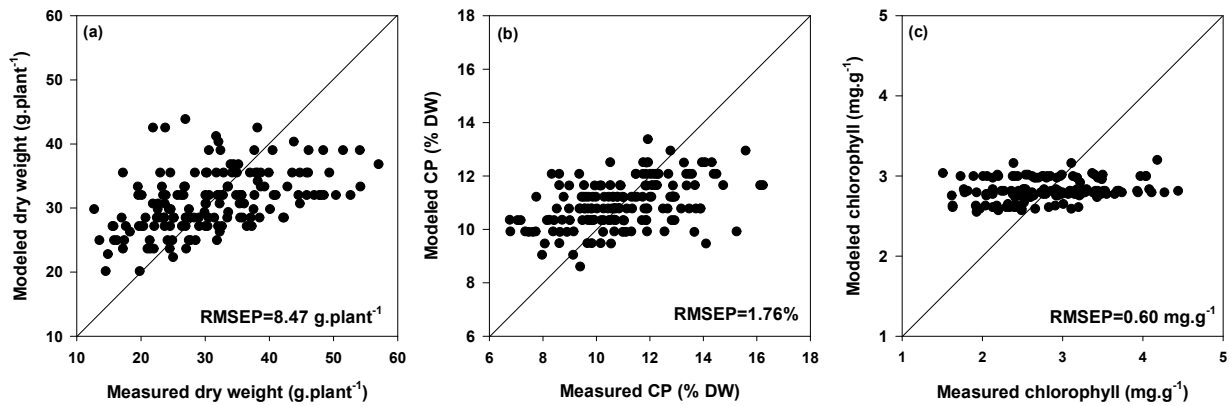
582

583

584 **Fig 1.** Modeled versus measured dry weight, nitrogen crude protein and chlorophyll content when  
585 fitting partial least square regression models to relate each biophysical characteristic to visual  
586 evaluations of biomass and greenness of 200 *Urochloa* hybrids.

587

588



589

590

591

592

593

594

595

596

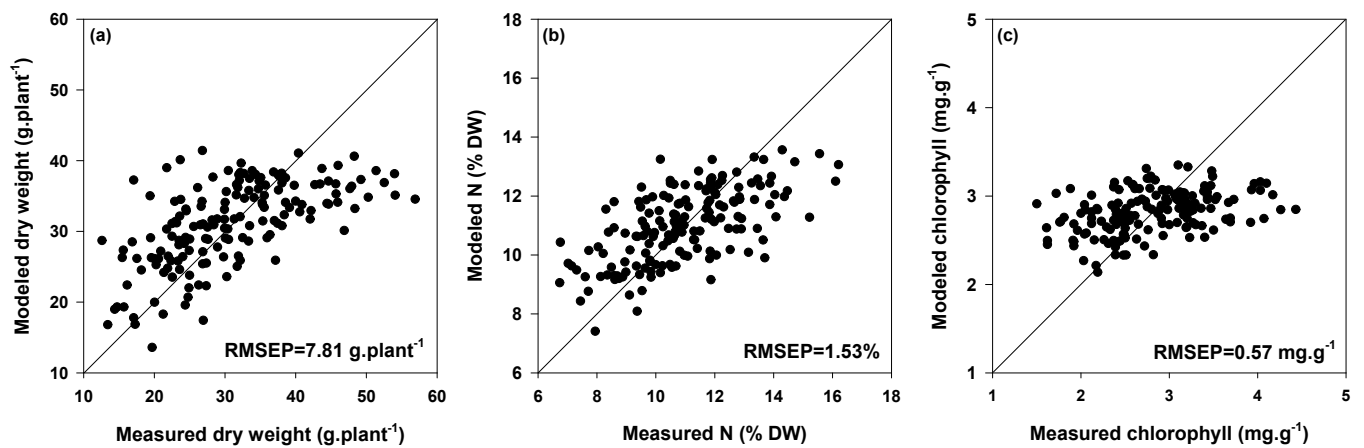
597

598

599

600

601 **Fig 2.** Modeled versus measured dry weight, nitrogen crude protein and chlorophyll content when  
602 fitting partial least square regression models to relate each biophysical characteristic to vegetation  
603 indices digital image analysis of 200 *Urochloa* hybrids.



607

608

609

610

611

612

613

614

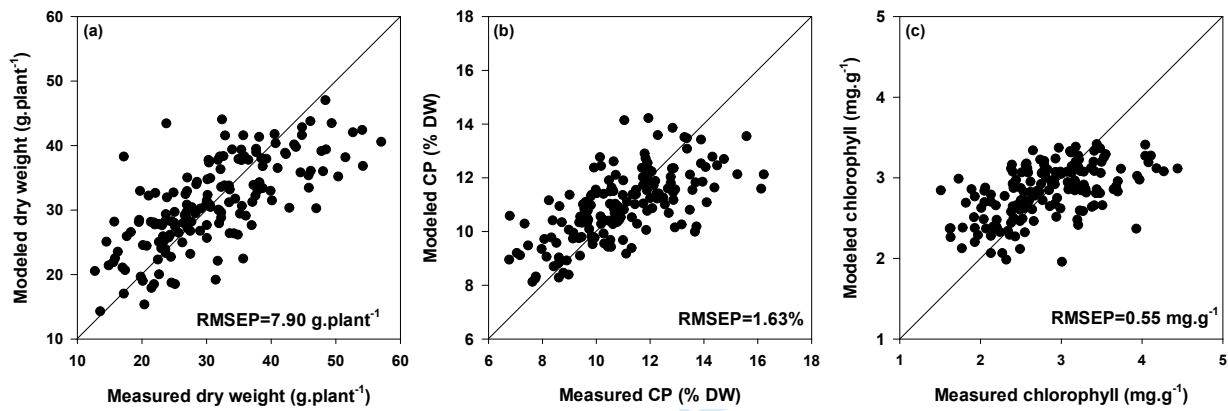
615

616 **Fig 3.** Modeled versus measured dry weight, nitrogen crude protein and chlorophyll content when  
617 fitting partial least square regression models to relate each biophysical characteristic to canopy  
618 spectral reflectance of 200 *Urochloa* hybrids.

619

620

621



622

623

624

625

626

627

628

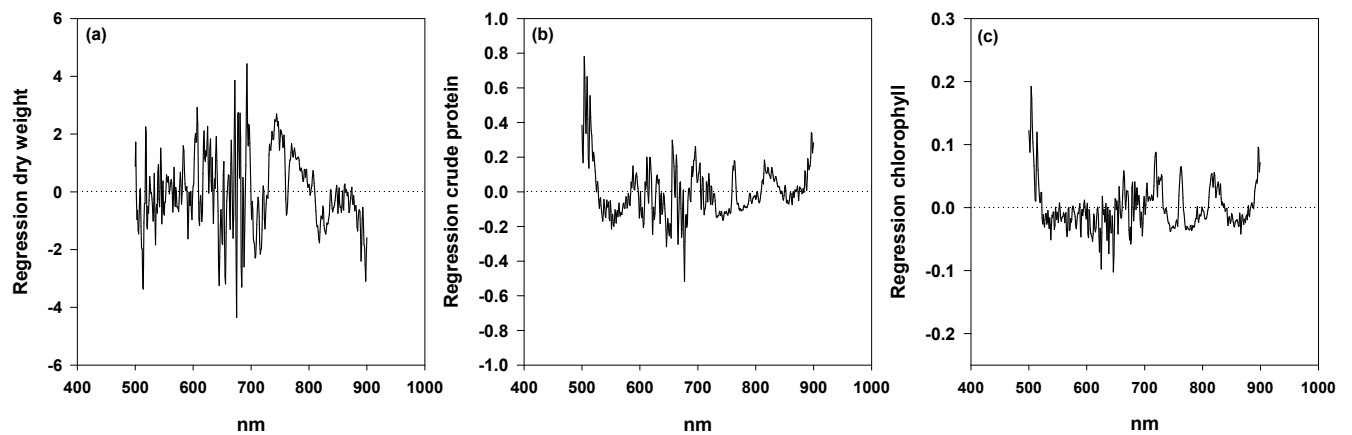
629

630



631 **Fig 4.** Regression coefficients of the fitted partial least squares regression models for dry weight,  
632 nitrogen and chlorophyll content crude protein and chlorophyll content. The regression coefficients  
633 represents the contribution of each spectral waveband to the overall prediction of each  
634 destructively-measured trait.

635



636

637

638

639

640

641

642

643

644

645

646

647

648

649 **Supplementary information**

650 **Supplementary Table 1.** Different protocols of spectral data collection and their respective root  
 651 mean squared error of prediction (RMSEP) for ~~nitrogen~~crude protein, dry weight and chlorophyll  
 652 content.

Collection	Trait	Factors¥	RMSEP
654 Day 1*	Dry weight (g plant <sup>-1</sup> )	4	9.23
	<del>Nitrogen</del> Crude protein (%)	11	1.29
	Chlorophyll (mg g FW)	4	0.49
655 Day 2*	Dry weight (g plant <sup>-1</sup> )	4	8.20
	<del>Nitrogen</del> Crude protein (%)	10	1.26
	Chlorophyll (mg g FW)	7	0.50
656 Day 3**	Dry weight (g plant <sup>-1</sup> )	4	7.63
	<del>Nitrogen</del> Crude protein (%)	11	2.07
	Chlorophyll (mg g FW)	5	0.54
657 Day 4**	Dry weight (g plant <sup>-1</sup> )	6	8.14
	<del>Nitrogen</del> Crude protein n (%)	2	1.21
	Chlorophyll (mg g FW)	3	0.58
659 All days	Dry weight (g plant <sup>-1</sup> )	6	7.90
	<del>Nitrogen</del> Crude protein (%)	5	1.63
	Chlorophyll (mg g FW)	5	0.55

660 Fifty plants were evaluated daily \* One scan collected per plant. \*\* Ten scans collected per plant.  
 661 ¥ Number of factors for which the root mean squared error of cross-validationprediction was  
 662 minimized in the model prediction.

663

664

665

666

667

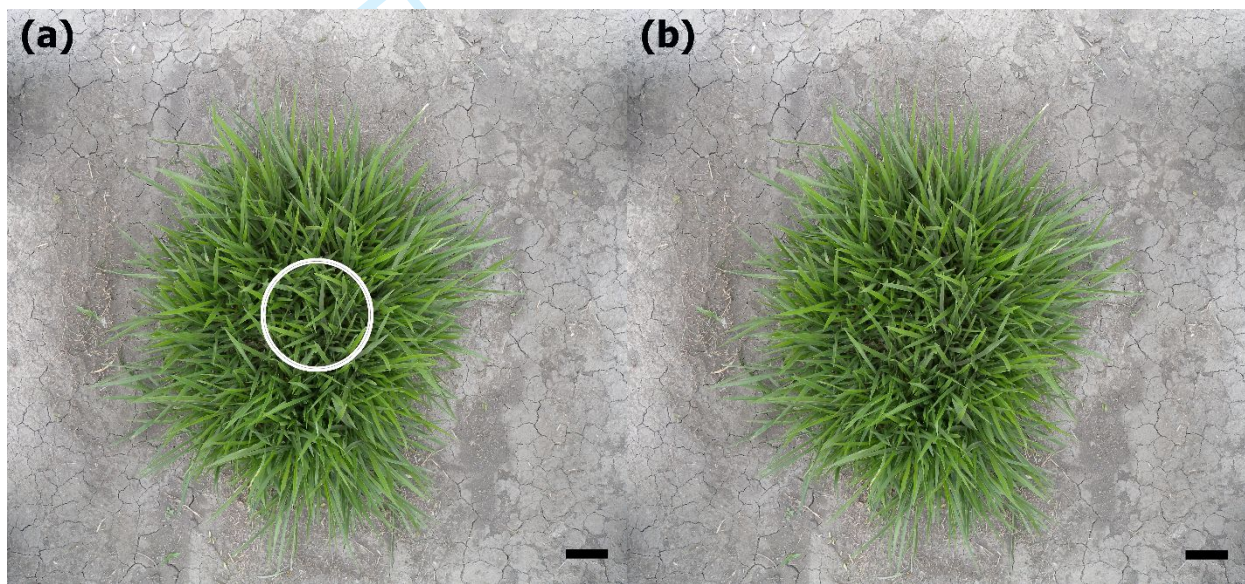
668

669

670

671

672 **Supplementary Fig 1.** Schematic representation of the observation geometry of hyperspectral  
673 analysis (a) and digital image analysis (b) techniques evaluated in 200 *Urochloa* hybrids. White  
674 circle positioned at the center of the plant canopy in figure (a) represents the 23-cm field of view  
675 of the spectroradiometer at a distance of 50mm from the plant canopy. For the digital image  
676 analysis (figure b), the whole plant, and not the 23-cm section, was used for segmentation and  
677 further analysis. Scale bar= 10 cm.



678

679

680

681

682

683

684

685

686

687

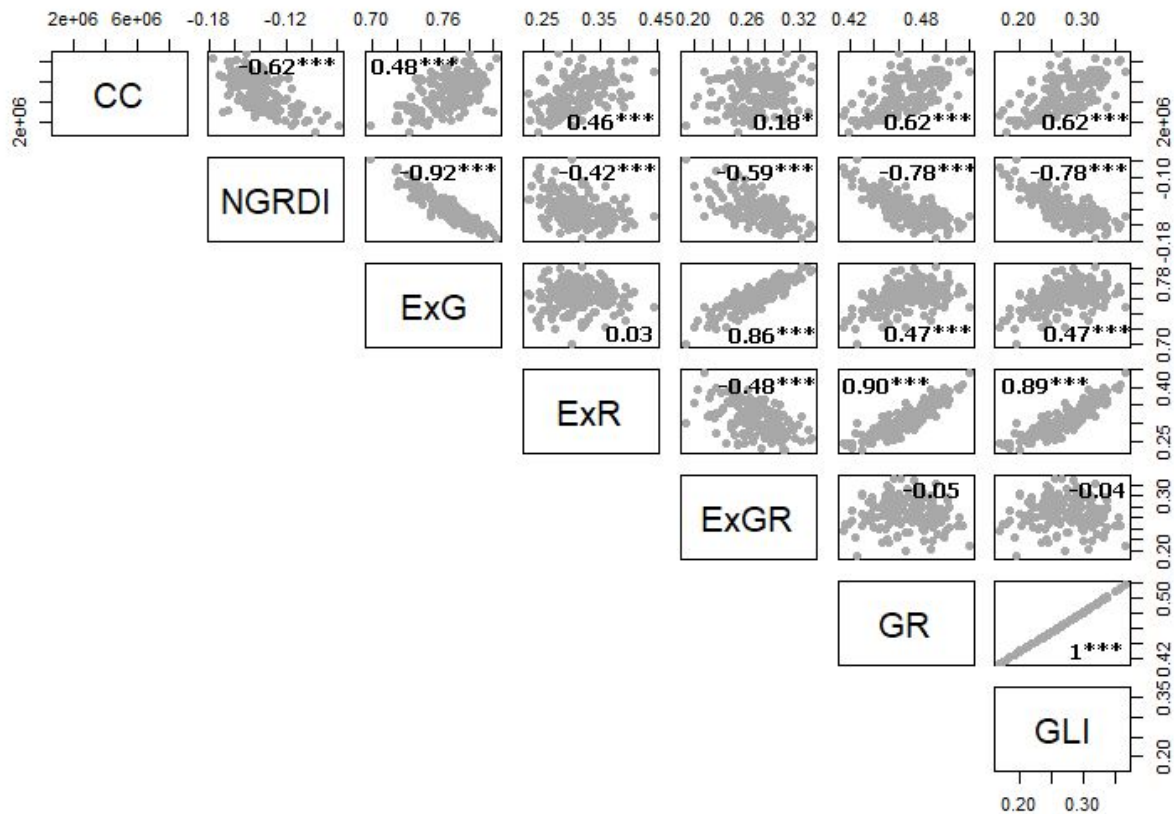
688 **Supplementary Fig 2.** Binary relationships and Pearson's correlation coefficients between seven689 plant traits extracted from digital images of 200 *Urochloa* hybrids. CC= canopy cover, NGRDI=

690 normalized green red difference index, ExG= excess green index, ExR= excess red index, ExGR=

691 excess green minus excess red, GR= green ration and GLI= green leaf index. Pearson's correlation

692 coefficients are indicated with their statistical significance as follows: \* $P \leq 0.1$ , \*\* $P \leq 0.01$ ,693 \*\*\* $P \leq 0.001$ .

694



695

696