Seed Science and Biotechnology ©2010 Global Science Books

Nutrient Homeostasis, C:N:S Ratios, Protein, and Oil Content in *Cuphea* Seed

Abdullah A. Jaradat* • Jana L. Rinke

United States Department of Agriculture (USDA) – Agricultural Research Service (ARS) 803 Iowa Avenue, Morris, MN 56267 USA

Corresponding author: * abdullah.jaradat@ars.usda.gov

ABSTRACT

Nutrient densities, carbon: nitrogen (C: N), nitrogen: sulphur (N: S), protein, and oil contents and their interrelationships were assessed during a 3-year study in seeds of the indeterminate *Cuphea* germplasm line PSR23, a potential oilseed crop selected from an inter-specific cross between *Cuphea lanceolata* and *C. viscosissima* of the Lythraceae. In order to mitigate the effect of indeterminacy on seed quality traits, the top 25% of biomass was removed at 100 and 200 growing degree days (GDD1 and GDD2, respectively) after flower initiation, as compared to no treatment (GDD0). Biochemical and nutrient densities were impacted by treatments and their interaction with years. Carbon and N allocations responded differently to treatments over years; nevertheless, large variances in C, and N contents and C: N ratio (R^2=0.70, 0.73, and 0.82, respectively), but not in N:S ratio (R^2=0.21) were explained by seed physical traits, treatments and their interaction with years. The C: N ratio and oil content, but not protein content, can be predicted as functions of nutrient densities in the seed with acceptable, but wide range of reliabilities. Removing the top 25% of biomass 100 GDD after flower initiation consistently resulted in increased oil content and in marked improvements in the reliability of its estimates. Oil and protein contents in *Cuphea* seed, being positively (r=0.83; p<0.01) and negatively (r=-0.86; p<0.01) correlated with the C: N ratio, respectively, may not be easily and concomitantly improved. N: S ratio, which is negatively correlated with total nutrient density in the seed, if reduced below 12: 1 could lead to larger oil content.

Keywords: C:N, *Cuphea*, indeterminacy, N:S, nutrients, oilseed, seed traits

Abbreviations: C, carbon; **CAN**, canonical discriminant function; **C.V.**, coefficient of variation; **GDD**, growing degree days; **N**, nitrogen; **NIPALS**, non-linear iterative partial least squares; **PC**, principal component; **PLSC**, partial least squares component; **REML**, restricted maximum likelihood

INTRODUCTION

The seed oil composition of *Cuphea lanceolata* and *C. viscosissima* (Lathyraceae) singled them out as candidates for domestication and for in-depth research. *C. lanceolata* is a protandrous, self-compatible, insect-pollinated allogamous diploid, with outcrossing rate >80% (2*x*=12), while *C. viscosissima* is a self-fertile, autogamous diploid (2*x*=12) and an outcrossing rate of about 30% (Hirsinger and Knowles 1984). A partial seed retention selection (PSR23) from a cross between *C. lanceolata* and *C. viscosissima* (Knapp and Crane 2000), referred to hereafter as *Cuphea*, is a potential new oilseed crop; its main fatty acids (i.e., capric, lauric, and myristic) have many potential uses in the detergent, lubricant, cosmetic and confectionary industries (Cermak and Isbell 2004).

Oil is the economically valuable component of *Cuphea* seed; however, protein in the seed meal is of potential value as animal feed (Evangelista *et al*. 2006). These components are inherited as quantitative traits and are influenced by the environment (Fageria 2001; Variath *et al*. 2009). Oil content in oilseed plants depends, among other factors, on seed weight, which in turn, depends not only on the source of assimilates during the seed-filling period but also on total mass of nutrients per seed (Henry and Westoby 2001), and on seed number per plant (Ruiz and Maddonni 2006). Henry and Westoby (2001) demonstrated that allocations of dry mass and nutrients are often not strictly proportional and suggested that a portion of the variation in seed mass may be explained by expenditure of nutrients.

Oilseed consists of N-free structural material, stored proteins, stored oils, and mineral nutrients (Fismes *et al*. 2000). The proportion of N-free structural material is expected to decrease with increasing seed weight, while protein and oil may compete for the remaining space in seeds. Oil and protein are negatively correlated in oilseed crops (Pipolo *et al*. 2004); however, this relationship is not a simple trade-off in the energy requirements for protein and oil synthesis (Bay and Becker 2004) and can be influenced by seed developmental stage (Variath *et al*. 2009). Li *et al*. (2006) concluded that more than 90% of seed N is in protein and ~50% of seed C is in oil; therefore, they suggested using C and N analysis as a first stage screening for oil content, especially in new oilseed crops or when seed sample size is limiting.

The vegetative growth phase of crops growing under short-season climates is of great importance for their seed yield and seed quality (Vilela *et al*. 2008). However, the indeterminate growth habit and the continued production of flowers throughout most of the vegetative growth phase may decrease the duration of the grain filling period (Ghasemnezhad and Honermeier 2007). In addition, the supra-optimal leaf area indices developed by *Cuphea* (Gesch *et al*. 2005), similar to other indeterminate oil crops (Khan and Lone 2005), may result in self shading and shading by other leaves within the highly-branched plant axis. Therefore, defoliation (Yang and Midmore 2004) or repeated biomass removal (Singer and Meek 2004) may impact the source-sink dynamics and, eventually, seed yield and seed quality of indeterminate plants.

We postulated that (1) the ratio between C and N (the most common nutrient elements in provisioned seed), and oil and protein contents, and their interrelationship are impacted by a number of seed-related traits such as average

Received: 8 February, 2010. Accepted: 25 February, 2010.

Original Research Paper

seed weight, number of seed per capsule, nutrient density (µg g^{-1} dry weight), and the seed packaging cost (i.e., ratio of capsule tissue to seed weight, and (2) timing of biomass removal may improve seed yield and quality of *Cuphea* through its impact on carbon allocation. The objective of this study was to quantify the impact of biomass removal on seed characteristics and to model the interrelationships between nutrients, C: N: S ratios, protein, and oil contents of *Cuphea* seed in an effort to improve its oil content.

MATERIALS AND METHODS

Field experiments

Experimental plots, arranged in a randomized complete block design with four replicates, were established in 2004-2006 on a Barnes-Buse loam (Barnes fine-loamy, mixed Udic Haploboroll, Buse fine-loamy, mixed, Udorthentic Haploboroll) at the Swan Lake Research Farm located near Morris, MN (45° 41′ N, 95° 48′ W, elevation 370 m). The field site was previously in corn, soybean, and corn for the 2004, 2005 and 2006 experiments, respectively. Planting (14 kg seed ha^{-1}; May 19, 2004; May 17, 2005; and May 8, 2006) and fertilizer application (110, 12 and 30 kg ha^{-1} of N, P and K, respectively) were done mechanically at a depth of about 15 mm. Seed produced in 2003, 2004, and 2005 were used for the 2004, 2005, and 2006 field experiments, respectively. Each plot consisted of six rows (6 m long and 60 cm row spacing); the middle two rows were mechanically harvested for seed yield determination at physiological maturity of seed in each treatment and year (Berti and Johnson 2008). In order to test whether biomass removal may improve seed yield and quality of *Cuphea* through its impact on carbon allocation in this indeterminate plant, all plants within each replicate and year, except the control, were subjected to source-sink manipulation by removing the top 25% of plant foliage at 100 and 200 growing degree days (GDD1 and GDD2, respectively) after flower initiation, as compared to no treatment (GDD0). Capsule and seed sampling for physical and chemical analyses followed the procedures described by Jaradat and Rinke (2008).

Determination of macro- and micro-nutrients

Seed samples were dried at 45°C in a forced air oven for a one week period or until no further reduction in weight occurred. Seed materials were ground in a coffee grinder and placed through a 1 mm screen (Thomas Scientific, NJ, USA). Carbon and nitrogen were determined on seed sub-samples using a LECO FP-428 analyzer (LECO, St. Joseph, MI), then the C: N ratio was calculated. Nitrogen was converted to protein by multiplying by 6.25. Digestion of plant materials for macro- and micro-nutrient determination followed the US-EPA 5051 method; this procedure was adapted using the Mars Xpress Microwave System from CEM (CEM Corp., NC, USA) sample preparation note XprAG-1. This microwave procedure uses 55 ml Teflon tubes in a 40 unit carousel. A 0.5 g sample weight was digested with 6.5 ml nitric acid (70% TMA, trace metal analysis) using a 15 min ramp program set to a power maximum of 1200 W and held for 15 min. The samples were allowed to cool to room temperature and transferred to 50 ml volumetric flasks and taken to volume with Milli-Q water (Millipore Corp., Billerica, MA, USA). Smaller samples were taken to 25 ml with adjustments made for HNO_3 concentrations. Analysis was completed using the Varian Vista-Pro CCD (charge coupled device, Varian Inc., Palo Alto, CA, USA) Simultaneous ICP-OES (inductively coupled plasma-optical emission spectroscopy) instrument. MNUSDA-STD 1-A and MNUSDA-STD 2 (Inorganic Ventures, Lakewood, NJ, USA) were prepared as elemental standards.

Determination of oil content

Seed sub-samples from each treatment, replicate and cropping season were used to estimate oil concentration according to established AOCS procedures (Bhardwaj *et al*. 2004), and summarized as follows: Oil was extracted from 2 g of ground seed at room temperature by homogenization for 2 min in 10 mL hexane/isopropanol (3: 2, vol/vol) with a Biospec Model 985-370 Tissue Homogenizer (Biospec Products, Inc., Racine, WI) and centrifuged at 4,000 × *g* for 5 min. The hexane-lipid layer was washed and separated from the combined extract by shaking and centrifugation with 10 mL of 1% $CaCl_2$ and 1% NaCl in 50% methanol. The washing procedure was repeated, and the purified lipid layer was removed by aspiration and dried over anhydrous Na_2SO_4. The oil content (g/100 g dry basis) was determined gravimetrically after drying in a vacuum oven at 40°C.

Statistical analyses

Data collected during 2004-2006 cropping seasons were tested for homogeneity of variances before conducting multivariate statistical analyses. Shapiro-Wilk's test was conducted to test for normality of the distribution of each variable and, if needed, to transform data in order to satisfy univariate and multivariate analysis of variance assumptions. Descriptive statistics (mean and coefficient of variation) were estimated for each variable and treatment (**Table 1**), and growing degree days were calculated with a base temperature of 10°C. Nutrient densities and growing degree days were log-transformed for statistical analyses (Zar 1996).

Canonical correlation analysis was performed to determine multi-trait associations at each year-GDD combination. Mixed models, using the restricted maximum likelihood method (REML), were constructed to quantify the impact of years, treatments, and covariates (seeds per capsule, seed weight and packaging cost) on total variance in each dependent variable (oil%, protein%, C: N ratio, N: S ratio, and nutrient densities in *Cuphea* seed). A whole model R^2 was calculated for each dependent variable and was partitioned according to its sources of variation (Payne *et al*. 2007). The partial least squares (PLS) regression option in the non-linear iterative partial least squares (NIPALS) algorithm (Esbensen 2005; Camo ASA 2007) was used on the transformed data to construct a set of components that accounts for as much variation as possible while modelling C: N ratio and oil content as functions of treatments, years, nutrient densities, and seed covariates. The PLS models developed in this analysis were cross-validated by successively leaving out data one at a time. A model was built using the remaining data points then the model created was used to predict the dependent variable (Esbensen 2005). Principal components analysis, a dimension reduction and perceptual mapping statistical procedure (StatSoft Inc. 2010), was employed to reduce the dimensionality of a matrix based on all factors (years and treatments) and variables (C: N ratio, N: S ratio, nutrient densities, protein%, and oil%) in the data set and to test for meaningful association among these factors and variables, and whether a globally optimum C: N provisioning ratio is impacted by allocation of other nutrient minerals to *Cuphea* seed, especially sulphur.

RESULTS

Descriptive statistics

Mean and coefficient of variation (CV) of physical and chemical properties of seed of *Cuphea* plants subjected to biomass removal treatments during early reproductive stage averaged over three growing seasons (**Table 1**) indicated that the treatments affected most traits irrespective of the trait level of variation (CV). The biomass removal treatments resulted in significantly larger, but lighter, number of seeds per capsule and a larger nutrient density per unit dry matter as compared to the control (GDD0). These differences were associated with significant changes in C: N ratio, and oil content, but not in packaging cost, N: S ratio or protein content. Oil content averaged 296 g kg^{-1} for the control and increased significantly (364 g kg^{-1}) due to the GDD1 treatment; however, the late treatment (GDD2) resulted in a significant drop in oil content (225 G KG^{-1}). Most traits displayed large levels of variation (CV), with the exception of C: N ratio and protein content. The slope of the relationship between oil and protein contents decreased gradually from 0.41 for GDD0 to 0.36 for GDD1, and finally to 0.25 for GDD2.

Table 1 Mean and coefficient of variation (C.V.) of physical and chemical properties of seed of *Cuphea* subjected to biomass removal during early reproductive stage averaged over three growing seasons. Treatments consisted of removal of the top 25% of plant foliage at 100 and 200 growing degree days, (GDD1and GDD2, respectively) after flower initiation as compared to a control (GDD0).

Variable	Treatments					
	GDD0		GDD1		GDD2	
	Mean	C.V.	Mean	C.V.	Mean	C.V.
Seeds/capsule	9.8 b	33.0	11.0 a	33.1	12.2 a	27.6
Seed weight g/100 seed	2.9 a†	25.7	2.6 b	29.8	2.3 c	34.5
Packaging cost	0.4	34.3	0.43	37.1	0.5	33.2
Nutrients µg g^{-1}	82.7 b	30.8	84.6 b	56.1	94.0 a	52.4
C:N	16.4 b	2.2	18.7 a	1.8	15.8 b	1.1
N:S	12.5	15.3	12.3	26.5	12.9	25.8
Oil g kg^{-1}	296 b	9.8	364 a	5.4	225 c	8.6
Protein g kg^{-1}	215	2.5	194	1.67	209	1.8
Oil/protein slope of regression model	0.41		0.36		0.25	

† Means of each variable followed by different letters differ significantly (Tukey HSD, 0.05)

Multivariate relationships

Two principal components, derived from all factors (treatments and years) and variables (oil content, protein content, nutrient density, C: N and N: S ratios), detected the underlying structure of and accounted for 67% of total variation in the data set (**Fig. 1A**). The first PC accounted for 47% of total variance and was dominated by differences between GDD2 and protein content on the positive side, and by oil content, C: N ratio, GDD1 and GDD0, in decreasing order, on the negative side. Differences among years (2004 and 2005) and N: S ratio on the positive side, and 2006 and nutrient density on the negative side dominated PC2 and explained 20% of total variation.

The standardized biplot (**Fig. 1B**) representing biochemical variables and samples of *Cuphea* seed together in two dimensions indicated that oil content, protein content and C: N ratio are associated with PC1 and contributed to the 47% variation explained by the principle component. Increasing values of both oil content and C: N ratio are associated with decreasing values of protein content on this component. Nutrient density and N: S ratio were closely associated, inversely related, and contributed to the 20% of total variation explained by PC2. The orientation of the variables suggested that plants subjected to GDD2 tend to have larger protein content, but smaller oil content and C: N ratio; whereas those subjected to GDD1 have larger oil content and C:N ratio, but smaller protein content. The inner and outer polygons contain 50 and 75% of the data points; minimum and maximum values of all five biochemical variables are indicated by the boundaries of the outer polygon.

Sources of variation in seed nutrients and biochemical contents

A highly complex picture emerged from the single and joint effects of years, treatments and covariates on the biochemical contents and nutrient densities in seed of *Cuphea* plants subjected to biomass removal treatments (**Table 2**). The variance portion, explained by all factors, in the biochemical contents was small for N: S ratio (0.21), intermediate for oil content (0.55) and large for protein content (0.73) and C: N ratio (0.82). Annual differences, unlike treatments and their interaction with years, did not contribute to any significant changes in these traits. All three covariates affected one or more of the biochemical traits; however, none of these traits was significantly impacted by all three covariates. Variation in number of seed per capsule resulted in significant differences in N: S ratio; variation in seed weight resulted in significant differences in protein content, C:N and N:S ratios; whereas, variation in packaging cost resulted in significant differences in oil and protein contents.

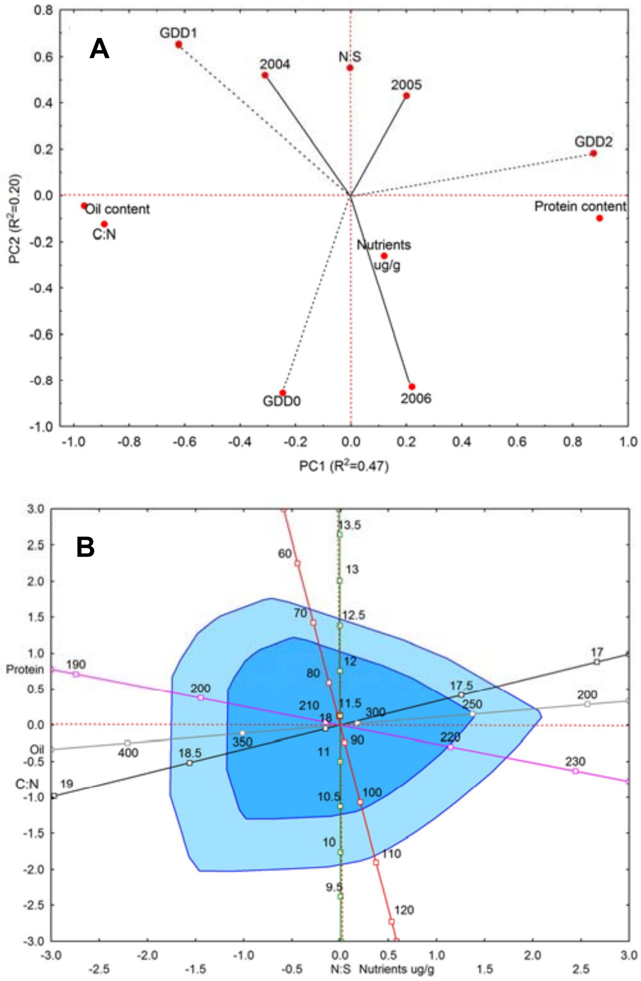

Fig. 1 (**A**) Variance accounted for by, and loadings of factors and variables on, the first two principal components (PC); and (**B**) standardized biplot of seed samples and their biochemical traits in seed of *Cuphea* subjected to biomass removal at the early reproductive stage during three cropping seasons.

The variance portion, explained by all factors, in individual nutrients was small for Fe (0.18), K (0.15), and Zn (0.12); intermediate for B (0.40), Ca (0.35), Cu (0.28), Mg (0.42), Mn (0.25), P (0.46), S (0.25) and Se (0.26); and large for Ba (0.75), C (0.70), and N (0.73); whereas the respective value for total nutrients density was 0.58. Treatments or their interaction with years accounted for significant portion of variation in most nutrient densities; however, treatments and their interaction with years accounted for significant portion of variation in B, N, and total nutrient density. None of the 14 nutrient densities was significantly impacted by variation in all three covariates. Variation in number of seeds per capsule significantly impacted the largest number of nutrients (11), followed, in decreasing order by variation in packaging cost (5), and seed weight (3). Total nutrient density was significantly impacted by all factors and covariates except years.

Relationships among seed nutrients and biochemical content

The first canonical discriminant function (CAN; **Fig. 2**) derived from seed biochemical contents explained 51% of variation in these traits with positive loadings of C: N ratio (0.77) and oil content (0.92) and negative loading of protein content (-0.58). The first CAN derived from nutrient densities explained 24% of their variation, with all nutrients, except Ba, having negative loadings. A strong and positive canonical correlation (r=0.91; p<0.0001) was found between these two sets of variables. Most seed samples representing

Table 2 Model fit (R^2) and test statistics of a mixed model ANOVA for biochemical contents and nutrient densities in seed of *Cuphea* seed subjected to biomass removal at the early reproductive stage during three cropping seasons.

Variable	R^2	Fixed factor	Random factors		Covariates		
		Treatment	Year	Year x Treatment	Seeds/ capsule	Seed weight	Packaging cost
			------------------------------Probability of a larger------------------------------				
		F-value	---------------z-value------------------		-----------------F-value---------------		
Oil	0.55	0.002	0.2	0.05	0.3	0.5	0.01
Protein	0.73	0.06	0.2	0.05	0.9	0.02	0.05
C:N	0.82	0.1	0.4	0.05	0.4	0.01	0.2
N:S	0.21	0.06	0.2	0.10	0.009	0.06	0.5
Boron	0.40	0.01	0.3	0.05	0.001	0.6	0.9
Barium	0.75	0.2	0.2	0.05	0.001	0.05	0.9
Calcium	0.35	0.2	0.9	0.05	0.001	0.9	0.3
Carbon	0.70	0.4	0.2	0.05	0.3	0.9	0.3
Cupper	0.28	0.05	0.1	0.20	0.001	0.3	0.6
Iron	0.18	0.1	0.1	0.10	0.001	0.2	0.3
Potassium	0.15	0.3	0.9	0.05	0.002	0.8	0.05
Magnesium	0.42	0.3	0.2	0.10	0.001	0.6	0.07
Manganese	0.25	0.03	0.3	0.10	0.001	0.01	0.5
Nitrogen	0.73	0.06	0.2	0.05	0.9	0.02	0.05
Phosphorus	0.46	0.3	0.2	0.10	0.008	0.7	0.04
Sulphur	0.25	0.4	0.2	0.10	0.005	0.6	0.07
Selenium	0.26	0.6	0.2	0.10	0.3	0.5	0.7
Zinc	0.12	0.5	0.9	0.05	0.002	0.5	0.4
Nutrients µg/g	0.58	0.06	0.1	0.05	0.0001	0.0001	0.0001

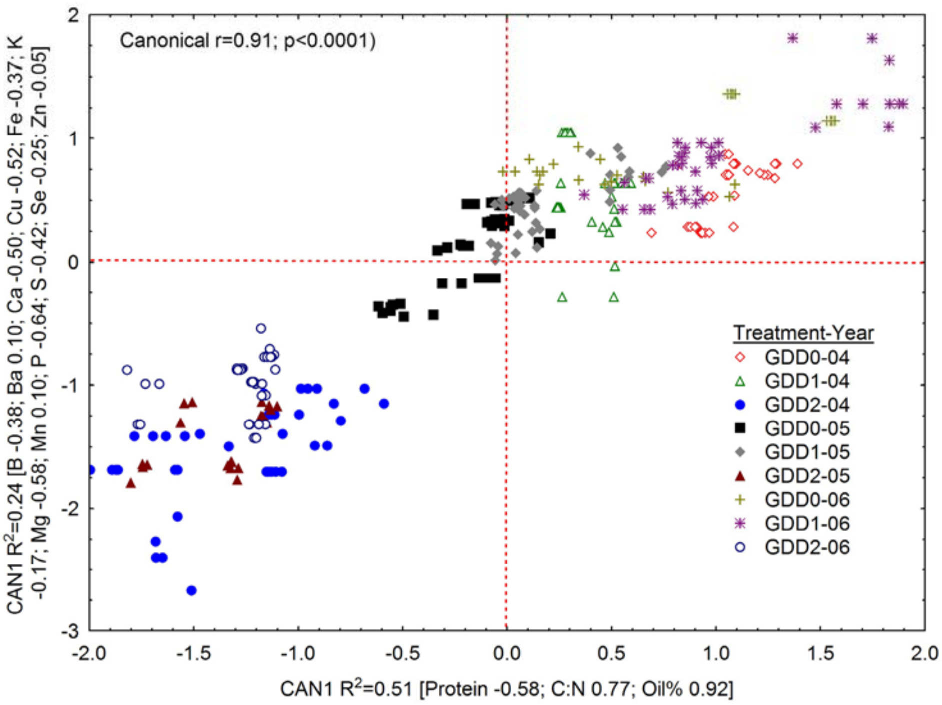

Fig. 2 Scatter plot and canonical correlation between biochemical contents and nutrient densities in seed of *Cuphea* subjected to biomass removal at the early reproductive stage during three cropping seasons.

GDD0 and GDD1 treatments have large and positive loadings on both discriminant functions; whereas those representing GDD2 have negative loadings on both discriminant functions. However, seed samples representing the GDD0 and GDD1 in 2005 (GDD0-05 and GDD1-05, respectively) clustered around the 0.0 coordinates of both discriminant functions. Variation among seed samples representing each treatment-year combination ranged from small (e.g., GDD0-04) to large (e.g., GDD2-04), especially along the first canonical discriminant function which was derived from nutrient densities.

Determinants and modelling of C: N ratio

Dynamic relationships among C, N and C: N ratio were displayed in response to different treatments during all three years (**Table 3**). Negative and significant correlations between N and each of C and C: N ratio were found in all years; however, the magnitude of these associations for GDD-0 was larger (r=-0.92) in 2004 as compared to 2005 (r=-0.25) or 2006 (r=-0.26). Due to treatments, the strength of the association (quantified by the correlation coefficient) between N and C decreased from -0.92 to -0.16 in 2004, but increased from -0.25 to -0.40 in 2005 and from -0.26 to -0.76 in 2006. Nevertheless, the negative and significant correlation between N and C: N ratio was maintained across all treatments and years. Similarly, the positive and significant correlation between C and C: N ratio was maintained across treatments and years; however, it decreased in magnitude from 0.73 to 0.34 in 2004, but increased from 0.47 to 0.91 in 2005, and from 0.69 to 0.91 in 2006.

The first and second partial least squared regression components (PLSCs) derived from all factors and covariates (listed in **Table 2**) captured 14 and 21% of the variance

Table 3 Simple correlation coefficients ($p<0.05$) between N, C, and C:N in seed of *Cuphea* subjected to biomass removal during early reproductive stage during three growing seasons.

Year	Element	Treatment					
		GDD0		GDD1		GDD2	
		Significant (p<0.05) correlation coefficients					
		C	C:N	C	C:N	C	C:N
2004	N	-0.92	-0.93	-0.63	-0.96	-0.16	-0.87
	C		0.73		0.55		0.34
2005	N	-0.25	-0.64	-0.21	-0.77	-0.40	-0.74
	C		0.47		0.58		0.91
2006	N	-0.26	-0.94	-0.86	-0.95	-0.76	-0.96
	C		0.69		0.81		0.91

in the predictors, respectively, and explained 51 and 10% of variation in the C: N ratio, respectively (**Fig. 3A**). The scatter plot and loadings of factors and covariates on the first two PLSCs indicated that the C: N ratio was impacted by annual differences, especially between 2004 and each of 2005 and 2006, as well as by differences between treatments and between covariates. The GDD1 treatment, especially in 2004 and 2006, resulted in larger C:N ratio as compared to GDD2, especially in 2005. The nutrients density, and to some extent, seed weight, had relatively small positive and negative impact on the C: N ratio estimates, respectively; whereas packaging cost and number of seeds per capsule had comparatively larger negative impacts.

A standardized biplot (**Fig. 3B**) representing both variables and samples of *Cuphea* seed together in two dimensions indicated that seeds per capsule and packaging cost are more closely associated together than they are associated with seed weight. All three variables, in increasing order, contributed to the 51% of variation explained in the C: N ratio by PLSC1. Nutrients density loaded high on PLSC2 and explained 10% of variation in the C: N ratio. The orientation of the variables suggested that increasing values of the first three variables are associated with decreasing values in nutrients density. The inner and outer polygons contain 50 and 75% of the data points; minimum and maximum values of all four variables are indicated by the boundaries of the outer polygon.

PLS regression models developed to validate the C: N ratio predictions for treatments or treatment-year combinations (**Table 4**) differed largely in their partial regression coefficients (based on log-transformed nutrient densities) and reliability as measured by RMSE and R^2 values. The PLS model developed for the whole data set explained 41% of variation in the C: N ratio and was associated with a large RMSE value, with only seven nutrients having significant partial correlation coefficients. Large values of B, Ba, Ca and Cu densities tended to cause a decrease in the C: N ratio; whereas large Fe, Se and Zn densities tended to

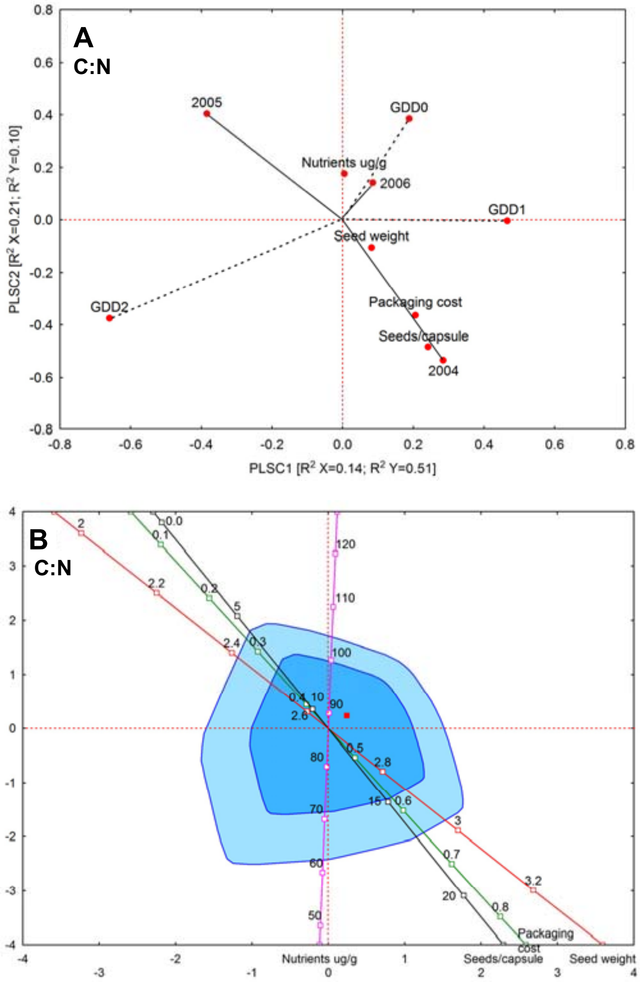

Fig. 3 (**A**) Variance in C: N ratio accounted for by, and loadings of factors and variables on, the first two partial least square regression components (PLSC); and (**B**) standardized biplot of seed samples and their physical and chemical traits as determinants of C:N ratio in seed of *Cuphea* subjected to biomass removal at the early reproductive stage during three cropping seasons.

increase it. Similarly, the PLS validation models developed for each treatment, averaged over years, explained 24, 43, and 50% of variation in the C: N ratio estimates for GDD0, GDD1 and GDD2, respectively. Large differences between GDD0 and each of GDD1 and GDD2, as well as between the last two treatments is evident in the number of individual nutrients and the magnitude and sign of their partial regression coefficients.

The treatment-year combination matrix represented a complex picture as to the number of positive or negative

Table 4 Partial Least Square (PLS) validation model statistics to predict C:N as a function of log-transformed nutrient densities in seed of *Cuphea* subjected to biomass removal at the early reproductive stage in three cropping seasons.

Treatment	Partial regression coefficient (p<0.05) Log(nutrient concentration) μg g^{-1}												Model	RMSE	R^2
	B	Ba	Ca	Cu	Fe	K	Mg	Mn	P	S	Se	Zn	βo		
All	-0.710	-0.136	-0.057	-0.148	0.194						0.100	0.189	16.7	0.300	0.41
GDD0	-0.286	0.347			0.289						0.226		10.0	0.330	0.24
GDD1		-0.178	-0.089	-0.171	0.367	-0.142	-0.036	-0.146		-0.074		-0.078	19.5	0.230	0.43
GDD2		0.097	0.081		0.041	0.054	0.032	0.064	0.054	0.053	0.038	0.061	12.8	0.100	0.50
GDD0-04	0.149	-0.090	0.024	0.031	0.034	0.020	0.066	0.108	0.148	0.061	0.015	0.090	13.3	0.021	0.80
GDD1-04	-0.157	0.027		-0.360	0.037	0.076	0.052	-0.056	0.080	0.040	-0.006	-0.023	16.2	0.033	0.92
GDD2-04						0.069	0.035	0.026	0.081	0.117			12.2	0.100	0.42
GDD0-05	0.020	-0.146	-0.017	-0.159	-0.021	0.004	0.024	-0.008	-0.010	0.027	-0.09	-0.010	17.1	0.060	0.62
GDD1-05	-0.020		-0.011	-0.027	-0.018	-0.011	0.010		0.010	-0.013	0.067	0.009	16.8	0.040	0.61
GDD2-05				-0.229			-0.022		0.039		0.513	-0.107	15.7	0.034	0.52
GDD0-06	-0.055	0.120	0.112	0.045	0.209	0.015	0.032	0.258	0.040	0.060	0.132	0.246	11.1	0.036	0.78
GDD1-06			-0.092		-0.506	-0.540	-0.680	-0.470	-0.480	-0.590	-0.950	1.070	38.4	0.110	0.87
GDD2-06					0.037		0.004	0.018	0.002	0.002	-0.012		15.8	0.022	0.58

significant partial regression coefficients, number and combination of nutrients contributing to the explained variance in the C: N ratio, and the reliability of the validation models as quantified by their RMSE and R^2 values. The C: N ratio of the control treatment (i.e., GDD0-04, GDD0-05, and GDD0-06) required all 12 nutrients to explain 80, 62, and 78% of its total variation, respectively. However, Ba and B were the only nutrients with negative impact on the C: N ratio estimates in GDD0-04 and GDD0-06, respectively; whereas eight nutrients had negative impacts on the C: N ratio estimates in GDD0-05. A slightly smaller number of nutrients was needed to explain the same or larger portion of variance in the C: N ratio estimates in GDD1 treatments over years. Eleven, ten and nine nutrients had significant partial regression coefficients in the validation models for the C: N ratio in GDD1-04, GDD1-05 and GDD1-06, respectively; the respective portion of variance in the C: N ratio explained by these models was 92, 61 and 87%. In addition, GDD1 treatments differed over years as to the magnitude and sign of the partial regression coefficients included in the validation models. During 2004, large densities of B, Cu, Mn, Se and Zn tended to decrease the C: N ratio; whereas, large densities of Ba, Fe, K, Mg, P, and S, tended to increase it. During 2005, large densities of six nutrients (B, Ca, Cu, Fe, K and S) tended to decrease C: N ratio estimates; whereas, large densities of the remaining four (Mg, P, Se and Zn) nutrients tended to increase it. Zinc was the only nutrient whose large density tended to increase the C: N ratio estimate during 2006; whereas, the remaining eight nutrients in the validation model tended to decrease it. The PLS models developed to validate C:N ratio estimates for the second treatment (GDD2) over years included a smaller number of nutrients with significant partial regression coefficients, and explained a smaller portion of variation in the C: N ratio. Five, five and six nutrients with significant partial regression coefficients were included in the PLS validation models for GDD2-04, GDD2-05 and GDD2-06, respectively; the respective portion of variance explained by these models was 42, 52, and 58%. Five nutrients (K, Mg, Mn, P and S) had positive impacts on the C: N ratio in GDD2-04; three nutrients (Cu, Mg, and Zn) had negative and two nutrients (P and Se) had a positive impact on the C: N ratio estimates in GDD2-05; whereas, four nutrients (Mg, Mn, P, and S) had a positive impact and one nutrient (Se) had a negative impact on the C: N ratio estimates in GDD2-06.

All nutrients, other than C and N, can be classified into three groups based on their significant negative impact on C:N ratio estimates. Each nutrient in the first group (Ba, K, Mg, P and S) contributed negatively to two treatment-year combinations. Each nutrient in the second group (B, Ca, Fe, Mn, and Zn) contributed negatively to three treatment-year combinations; whereas, the remaining nutrients (Cu and Se) contributed negatively to four treatment-year combinations. In addition, one nutrient (Ba) appeared in four regression models; two nutrients (B and Ca) each appeared in five regression models; one nutrient (Cu) appeared in six regression models; four nutrients (Fe, K, Mn, and Zn) each appeared in seven regression models; two nutrients (S and Se) each appeared in eight regression models; and two nutrients (Mg and P) each appeared in all nine regression models.

Determinants and modelling of oil content

The first and second PLSCs derived from all factors and covariates (listed in **Table 2**) captured 17 and 25% of the variance in the predictors, respectively, and explained 45 and 13% of variation in oil content, respectively (**Fig. 4A**). The scatter plot and loadings of factors and covariates on the first two PLSCs indicated that oil content was impacted by annual differences, especially between 2005 and 2006, as well as by differences between treatments (especially between GDD1 and GDD2) and between covariates (seed weight and each of nutrients density, packaging cost and number of seed per capsule); large values of the last three

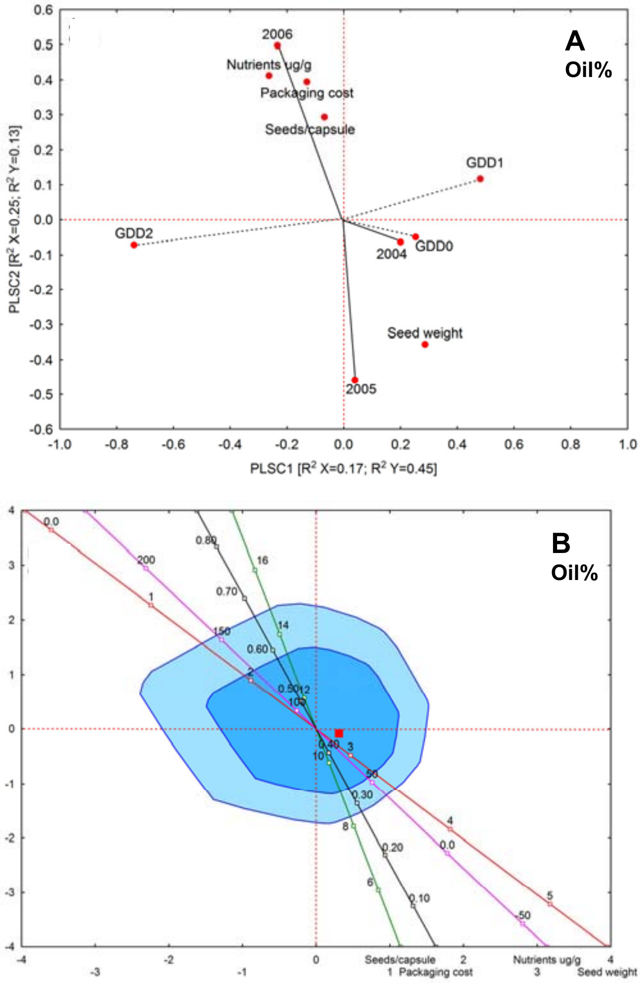

Fig. 4 (**A**) Variance in oil content accounted for by, and loadings of factors and variables on the first two partial least square regression components (PLSC); and (**B**) standardized biplot of seed samples and their physical and chemical traits as determinants of oil content in seed of *Cuphea* subjected to biomass removal at the early reproductive stage during three cropping seasons.

covariates tended to reduce oil content; whereas, heavier seed tended to contain more oil. The GDD0 and GDD1 treatments, especially in 2004 and 2005, resulted in larger oil content as compared to GDD2, especially in 2006.

The standardized biplot (**Fig. 4B**) representing both variables and samples of *Cuphea* seed together in two dimensions indicated that seeds per capsule and packaging cost are more closely associated with each other than with seed weight or nutrients density. All four variables (i.e., seed weight, nutrient density, packaging cost and number of seed per capsule), in decreasing order, contributed to the 58% of variation explained in oil content by both PLSCs. The orientation of the variables suggested that the heavier the seed, the smaller the nutrient density and the smaller the packaging cost. Also, the orientation of the variables for number of seeds per capsule and seed weight indicated a negative association between them. The inner and outer polygons contain 50 and 75% of the data points; minimum and maximum values of all four variables are indicated by the boundaries of the outer polygon.

The treatment-year combination in the PLS validation regression model for the whole data set indicated that 51% of total variation in oil content can be explained by the variation in only six nutrients (B, Ba, Ca, Cu,Se and Zn) whether N was used in the model or not (**Table 5**). The reliability of prediction (i.e., R^2 value) was smallest (25%) for GDD0, largest (78%) for GDD1 and intermediate (43%) for GDD2. Oil content in seed subjected to the control treatments in each year (i.e., GDD0-04, GDD0-05, and GDD0-

Table 5 Partial Least Square (PLS) validation model statistics to predict oil content as a function of log-transformed nutrient densities in seed of *Cuphea* subjected to biomass removal at the early reproductive stage in three cropping seasons.

Treatment	Partial regression coefficient ($p<0.05$) Log(nutrient concentration) μg g^{-1}												Model	RMSE	R^2
	B	Ba	Ca	Cu	Fe	K	Mg	Mn	P	S	Se	Zn	βo		
All	-7.620	-1.740	-1.040	-1.630							-1.540	1.510	56.85	3.04	0.51;0.51
GDD0	-1.105	1.855	0.289			0.360		1.470			-1.430	1.320	13.90	2.80	0.25;0.25
GDD1	1.154	4.752	0.901	-1.270	0.378	0.815	1.796	0.237	1.116	1.103	3.268	1.061	-42.25	0.62	0.78;0.78
GDD2					-0.644	0.368				0.446			17.90	0.50	0.43;0.44
GDD0-04	-0.259	0.137	-0.033	-0.100					-0.229		-0.024	-0.119	34.30	0.63	0.40;0.36
GDD1-04	-0.396	0.100		-0.910	0.153	0.224	0.160		0.232	0.128	-0.012		26.60	0.14	0.91;0.94
GDD2-04	-0.020		0.060	0.156	0.110	0.080	0.072					0.105	16.40	0.49	0.44;0.51
GDD0-05	0.167			-0.935							-0.446	-0.279	27.50	0.50	0.42;0.43
GDD1-05	-0.390		-0.242	-0.589	-0.308	-0.214	-0.205	-0.09	-0.192	-0.235		-0.245	41.20	0.26	0.89;0.88
GDD2-05	-6.540	0.104	0.064	0.158	-0.060	-0.047			0.020	0.401	0.401		40.60	0.08	0.42;0.39
GDD0-06	-0.554	1.240	1.010	0.297	3.070	0.065	0.341	2.530	0.421	0.510	-0.462	1.850	-24.34	0.87	0.94;0.94
GDD1-06	1.440	5.470	0.942	-1.430		0.437	1.427		0.835	0.717	2.857	2.181	-32.60	0.53	0.84;0.85
GDD2-06		-0.163			0.084							-0.198	19.30	0.11	0.20;0.21

06) required four, seven, and 12 nutrients, respectively, to explain the respective 40, 42, and 94% of its total variation. However, most nutrients had negative impact on oil content estimation in GDD0-04 and GDD0-05; whereas only two nutrients (B and Se) had negative impacts on oil content estimates in GDD0-06. A large number of nutrients was included in the PLS validation regression models and accounted for larger portions of variance in oil content estimates in GDD1 treatments over years. Nine, 10 and 10 nutrients had significant partial regression coefficients in the validation models for oil content in seed subjected to GDD1-04, GDD1-05 and GDD1-06 treatments, respectively; the respective portion of variance in oil content explained by these models was 91, 89 and 84%. In addition, the GDD1 treatments differed over years as to the magnitude and sign of the partial regression coefficients included in the validation regression models. During 2004, large densities of B, Cu, and Se tended to decrease oil content; whereas, large densities of Ba, Fe, K, Mg, P, and S, tended to increase it. During 2005, large densities of all nutrients, except Ba and Se, tended to decrease oil content estimates. Copper was the only nutrient whose large density tended to decrease oil content estimate during 2006; whereas the remaining nutrients in the validation regression model, except Fe and Mn, tended to decrease it. The PLS regression models developed to validate oil content estimates for the second treatment (GDD2) over years included smaller number of nutrients with significant partial regression coefficients, and explained smaller portion of variation in oil content. Seven, nine, and three nutrients with significant partial regression coefficients were included in the PLS validation regression models for GDD2-04, GDD2-05 and GDD2-06, respectively; the respective portion of variance explained by these models was 44, 42, and 20%. Boron was the only nutrient with a negative impact, while the remaining six nutrients (Ca, Cu, Fe, K, Mg, and Zn) had positive impacts on oil content in GDD2-04; three nutrients (B, Fe and K) had negative and six nutrients (Ba, Ca, Cu, P, S, and Se) had a positive impact on oil content estimates in GDD2-05; and only Fe had positive and Ba and Zn had a negative impact on oil content estimates in GDD2-06.

All 12 nutrients can be classified into five groups based on their significant negative impact on oil content estimates. Boron was the only nutrient to have negative impact in six out of nine PLS regression models; Cu in five PLS models; Se and Zn, each in four PLS models; Ca, Fe, K, and P, each in two PLS models; and Ba, Mg, Mn and S, each in one PLS model. Finally, N did not improve the reliability of prediction when included in each of the nine PLS models; the resulting R^2 values were similar or slightly larger than those obtained without it.

Oil content, based on its positive association with the C: N ratio for the whole data set, can be predicted by the linear regression model:

$$\text{Oil content}=97 * (\text{C: N}) - 1260;\ R^2=0.63,\ p<0.0001 \quad [1]$$

The reliability with which oil content can be predicted in seed of plants not subjected to treatment (i.e., GDD0) was large and can be expressed by the model:

$$\text{Oil content}=74 * (\text{C: N}) - 904;\ R^2=0.78,\ p<0.0001 \quad [2]$$

The reliability of prediction (R^2) was slightly reduced in seed of plants subjected to GDD1, where

$$\text{Oil content}=31 * (\text{C: N}) - 197;\ R^2=0.74,\ p<0.0001 \quad [3]$$

and it was drastically reduced in seed of plants subjected to GDD2

$$\text{Oil content}=45 * (\text{C: N}) - 516;\ R^2=0.43,\ p<0.001 \quad [4]$$

DISCUSSION

In *Cuphea*, number of flowers, fertile capsules, seeds per capsule and single seed weight are not established sequentially during plant development due to its indeterminate growth habit (Berti and Johnson 2008; Jaradat and Rinke 2008). Therefore, *Cuphea* plants may not achieve their yield potential due to inadequate seed development (Berti *et al*. 2007; Vilela *et al*. 2008). Consequently, seed developed late in the growing season may have small oil content as a result of the relative in-availability of resources (e.g., C and N) at the time of provisioning (Lalonde 1988; Berti *et al*. 2007).

McGinley and Charnov (1988) indicated that optimal seed mass, expressed as C allocation, is positively related to C:N ratio because the optimal investment of C-based compounds increases as C:N ratio increases. We found large within-capsule and within-plant variation in seed number and seed mass, respectively, in *Cuphea* (**Table 1**; Jaradat and Rinke 2008); when contrasted with the C:N ratio, the latter may be considered as an ultimate cause of within-plant seed mass variability as suggested by Mabry and Wayne (1997). The smaller seed weight and larger C: N ratio under late biomass removal (GDD2) could be due not only to increased packaging cost, but also to higher expenditure of nutrients per seed (**Table 1**) as suggested by Henry and Westoby (2001).

Removal of the top 25% of biomass at GDD1 enhanced plant growth and yield, increased seed number per plant and resulted in significantly larger oil content as compared to the control. Biomass thinning or defoliation treatments when imposed within the critical period for seed set, produced comparable results in *Brassica juncea* L. (Khan and Lone 2005), *Helianthus annuus* L. (Ruiz and Maddonni 2006), and *Abutilon theophrasti* (Mabry and Wayne 1997). However, simulated removal of the top 30% of biomass in soybean [*Glycine max* (L.) Merr.] at the mid vegetative to early reproductive stages (Singer and Meek 2004), resulted in slightly smaller grain yield and pod number m^{-2}, almost similar yield components, and similar protein and oil contents as compared to the control.

Biomass removal treatments caused positive changes in the physical structure of the *Cuphea* canopy (results not shown) and may have resulted in increased proportion of foliage receiving direct sunlight (Mabry and Wayne 1997). We speculate, based on Mabry and Wayne's results (1997), that plants subjected to GDD1 aborted fewer flowers and capsules throughout the growing season, compared to the control, due to increased canopy light penetration and transient increases in leaf-level photosynthetic rate. However, we still have to determine whether reproductive output in this indeterminate species is regulated by adjusting the number of flowers and capsules aborted or by adjusting the number of seeds per capsule or single seed weight. Li *et al.* (2006) reported larger seed size, seed yield and oil content, and a concomitant increase in the C: N ratio as a result of better light conditions in *Arabidopsis thaliana*. We suggest that, at least, the GDD1 treatment may have resulted in larger oil content due to better light penetration through the less dense foliage of *Cuphea* as it can attain a leaf area index of ~6 under normal field growing conditions (Gesch *et al.* 2005).

Oil content was reported by many authors (e.g., Leach *et al.* 1999 and references therein; Ruiz and Maddonni 2006) to be a conservative trait; however, *Cuphea* seed displayed large variation in oil content due to biomass removal treatments (**Table 1**) as well as to annual variation and its interaction with treatments (**Table 2**). Berti *et al.* (2007) suggested that *Cuphea* should be harvested after 265 GDD post anthesis when seed have attained maximum weight and oil content. However, Gesch *et al.* (2005) and Ghasemnezhad and Honermeier (2007) found that total seed oil content can be influenced by harvest date of *Cuphea* and *Oenothera biennis*, respectively. Nevertheless, Ruiz and Maddonni (2006) reported an equal seed weight change, but not seed oil concentration, per each change in the source-sink ratio in *Helianthus annuus* L.

Oil and protein synthesis are negatively correlated in *Cuphea* (**Figs. 1, 2**) as well as in most oilseed crops (Fismes *et al.* 2000; Fageria 2001). This may (Pipolo *et al.* 2004) or may not (Variath *et al.* 2009) constitute a significant barrier to simultaneously improving their contents. We found that the oil/protein slope of the regression model in this study decreased from 0.41 (GDD0) to 0.36 (GDD1) then to 0.25 (GDD2) as compared to 0.19 reported for *Glycine max* (L.) Merr. (Pipolo *et al.* 2004). These values are substantially less than the 0.82 value expected based on oil and protein relative synthesis per unit photosynthates according to Pipolo *et al.* (2004) who remarked that increasing the protein content of the seed did not result in nearly as large a decrease in oil content as expected based on the energy required for protein production. Obviously, this negative relationship between oil and protein contents is not a simple trade-off in the energy requirements for protein and oil synthesis (Variath *et al.* 2009).

The strong linear relationships established between C:N ratio and oil content (equations [1] to [4]) suggest that C and N analyses of seed can be used as an alternative to oil content estimation. Li *et al.* (2006) reported large plant-to-plant variation in oil content of *Arabidopsis thaliana* with a strong relationship with the C: N ratio (R^2=0.79; p<0.001) and emphasized its importance in screening for oil content in early segregating generations, or when seed is limited. *Cuphea* appears to adjust C and N resource allocation at the earliest stage of reproductive development after a single removal of the upper 25% biomass 100 GDD after flower initiation and to accumulate significantly larger oil content as compared to the control; however, we need to determine whether reproductive output in this indeterminate species is regulated by adjusting the number of flowers and capsules aborted or by adjusting the number of seeds per capsule or single seed weight.

CONCLUSIONS

Oil is the economically valuable component in seed of indeterminate *Cuphea* plants; however, its content depends, among other factors, on seed weight, which in turn, depends not only on the source of assimilates during the seed-filling period but also on total mass of nutrients per seed, and on seed number per plant. This is the first report on the relationships between densities of 14 nutrients, C: N and N: S ratios, and oil and protein contents assessed in seeds of indeterminate *Cuphea* plants subjected to biomass removal after flower initiation. Oil and protein contents in *Cuphea* seed, being positively and negatively correlated with the C: N ratio, respectively, may not be easily and concomitantly improved unless the currently large N: S ratio can be reduced through proper plant nutrition. Oil content can be improved by removing the top ~25% of plant foliage no later than 10-15 days after flower initiation to help synchronize flower production, allocations of dry mass and nutrients to the developing seed and seed maturation.

ACKNOWLEDGEMENTS

We thank Jay Hanson for C and N determination and Beth Burmeister for editing the manuscript. The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the United States Department of Agriculture or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable. USDA is an equal provider and employer.

REFERENCES

Baye T, Becker H (2004) Analyzing seed weight, fatty acid composition, oil and protein contents in *Vernonia galamensis* germplasm by near-infrared reflectance spectroscopy. *Journal of the American Oil Chemists' Society* **81**, 641-645

Berti M, Johnson B (2008) Growth and development of *Cuphea*. *Industrial Crops and Products* **27**, 265-271

Berti M, Johnson B, Manthey L (2007) Seed physiological maturity in *Cuphea*. *Industrial Crops and Products* **25**, 190-201

Bhardwaj H, Hamama A, van Santen E (2004) Fatty acid and oil content in white lupine seed as affected by production practices. *Journal of American Oil Chemists' Society* **81**, 1035-1038

Camo ASA (2007) *The Unscrambler software* v 9.7, Camo ASA, Oslo

Cermak S, Isbell T (2004) Synthesis and physical properties of *Cuphea* – oleic estolids and esters. *Journal of the American Oil Chemists' Society* **81**, 297-203

Esbensen K (2005) *Multivariate Data Analysis – in Practice* (5th Edn), CAMO Process AS, Oslo, Norway, 598 pp

Evangelista R, Wu Y, Hojilla-Evangelista M (2006) Characterization of proteins in *Cuphea* seed. *Journal of the American Oil Chemists' Society* **83**, 785-790

Fageria V (2001) Nutrient interactions in crop plants. *Journal of Plant Nutrition* **24**, 1269-1290

Fismes J, Vong P, Guckert, A, Frossard E (2000) Influence of sulphur on apparent N-use efficiency, yield and quality of oilseed rape (*Brassica napus* L.) grown on calcareous soil. *European Journal of Agronomy* **12**, 127-141

Gesch R, Cermak S, Isbell T, Forcella F (2005) Seed yield and oil content of *Cuphea* as affected by harvest date. *Agronomy Journal* **97**, 817-822

Ghasemnezhad A, Honermeier B (2007) Seed yield, oil and fatty acid composition of *Oenothera biennis* L. affected by harvest date and harvest method. *Industrial Crops and Products* **25**, 274-281

Henry M, Westoby M (2001) Seed mass and seed nutrient content as predictors of seed output variation between species. *Oikos* **92**, 479-490

Hirsinger F, Knowles P (1984) Morphological and agronomic descriptors of selected *Cuphea* germplasm. *Economic Botany* **38**, 439-451

Jaradat A, Rinke J (2008) Flowering, capsule and seed characteristics in *Cuphea*. *Euphytica* **161**, 447-459

Khan N, Lone P (2005) Effects of early and late season defoliation on photosynthesis, growth and yield of mustard (*Brassica juncea* L.). *Brazilian Journal of Plant Physiology* **17**, 181-186

Knapp S, Crane J (2000) Registration of reduced shattering *Cuphea* germplasm PSR23. *Crop Science* **41**, 299-300

Lalonde R (1988) A test of McGinley and Charnov's multiple resource pool hypothesis. *Evolutionary Ecology* **2**, 316-320

Leach J, Stevenson H, Rainbow A, Mullen L (1999) Effect of high plant populations on the growth and yield of winter oilseed rape (*Brassica napus*). *Journal of Agricultural Science* **132**, 173-180

Li Y, Beisson F, Pollard M, Ohlrogge J (2006) Oil content of Arabidopsis seed: the influence of seed anatomy, light and plant-to-plant variation. *Phytochemistry* **67**, 904-915

Mabry C, Wayne P (1997) Defoliation of the annual herb *Abutilon theophrasti*: mechanisms underlying reproductive compensation. *Oecologia* **111**, 225-232

McGinley M, Charnov E (1988) Multiple resources and the optimal balance between size and number of offspring. *Evolutionary Ecology* **2**, 77-84

Payne R, Harding S, Murray D, Soutar D, Baird D, Welham S, Kane A, Gilmour A, Thompson R, Webster R, Tunnicliffe G (2007) *The Guide to GenStat Release 10, Part 2: Statistics*. VSN International, Hemel Hempstead.

Pipolo A, Sinclair T, Camara G (2004) Protein and oil concentration of soybean seed cultured *in vitro* using nutrient solutions of differing glutamine concentration. *Annals of Applied Biology* **144**, 223-227

Ruiz R, Maddonni G (2006) Sunflower seed weight and oil concentration under different post-flowering source-sink ratios. *Crop Science* **46**, 671-680

Singer J, Meek D (2004) Repeated biomass removal affects soybean resource utilization and yield. *Agronomy Journal* **96**, 1382-1389

StatSoft Inc. (2010) STATISTICA (data analysis software systems) Version 9.1 www.statsoft.com

Variath M, Wu J, Li Y, Chen G (2009) Genetic analysis for oil and protein contents of rapeseed (*Brassica napus* L.) at different developmental times. *Euphytica* **166**, 145-153

Vilela A, Gonzalez-Paleo L, Rondanini D, Ravetta D (2008) Biomass allocation patterns and reproductive output of four *Oenothera* L. accessions native to Argentina. *Industrial Crops and Products* **27**, 249-256

Yang Z, Midmore D (2004) Experimental assessment of the impact of defoliation on growth and production of water-stressed maize and cotton plants. *Experimental Agriculture* **40**, 189-199

Zar J (1996) *Biostatistical Analysis* (3rd Edn), Prentice Hall, NJ, 662 pp