Mapping of the Egusi Seed Trait Locus (eg) and Quantitative Trait Loci Associated with Seed Oil Percentage in Watermelon

Jason Prothro, Katherine Sandlin, and Rattandeep Gill
Institute of Plant Breeding, Genetics and Genomics, University of Georgia, 111 Riverbend Road, Athens, GA 30602

Eleni Bachlava, Victoria White, and Steven J. Knapp
Monsanto, Woodland, CA 95695

Cecilia McGregor1
Department of Horticulture and Institute of Plant Breeding, Genetics and Genomics, University of Georgia, 1111 Miller Plant Sciences, Athens, GA 30602

ADDITIONAL INDEX WORDS. CIM, MIM, epistasis, QTL

ABSTRACT. The egusi watermelon (Citrullus lanatus) is popular in West Africa for its oil and protein-rich seed, which is consumed in soups and stews. The egusi phenotypic trait is controlled by a single recessive gene (eg) and is characterized by large seed size and fleshy, thick pericarp. An F2 mapping population was derived from Strain II (PI 279461) of the Japanese cultivar Yamato-cream with normal seed type and low seed oil percentage (SOP = 25.2%) and an egusi type from Nigeria [Egusi (PI 560023)] with high SOP (40.6%). Genetic analysis confirmed that the egusi seed trait is controlled by a single recessive gene (eg) and the location of the gene was mapped to 57.8 cM on linkage group (LG) 2, between markers NW0248325 and NW0250248. Four main quantitative trait loci (M-QTL) were identified for SOP in the population with the eg locus contributing 84% of the explained phenotypic variation (R²). A significant epistatic interaction (E-QTL) was identified between the eg locus and an M-QTL on LG 9B. The present study reports the location of the eg locus responsible for the egusi seed trait in watermelon on LG 2 as well as M-QTL and E-QTL associated with SOP.

Egusi seed is a part of the daily diet in many West African countries, including Ghana, Nigeria, and Benin [National Research Council of the National Academies (NRC), 2006]. Although the term “egusi” can be used to describe a certain seed type from several species of the Cucurbitaceae family (Achigan-Dako et al., 2008; NRC, 2006), it is often used in reference to egusi watermelon. The plants cultivated for their oil and protein-rich seeds are currently classified as C. lanatus ssp. mucospermus var. egusi (Fursa, 1972; Jeffrey, 2001) and have sometimes been misclassified as C. colocynthis (Gusmini et al., 2004), but the nature of the gene and its origin is unknown. Various approaches can be followed to determine the nature of a gene, including positional cloning, which requires knowledge of the location of a gene in the genome. Until recently the genetic resources needed for such an undertaking were not available in watermelon. However, recently a single nucleotide polymorphism (SNP) map was developed for an F2 population segregating for the egusi seed phenotype as well as SOP (Prothro, 2010; Sandlin et al., 2012). We used this resource to 1) determine the genomic location of the eg locus in watermelon; and 2) identify M-QTL and E-QTL associated with SOP in a population segregating for the egusi seed trait.

Materials and Methods

Seed of watermelon accessions and cultivars with normal or egusi seed types were obtained from the U.S. Department of Agriculture–Agricultural Research Service germplasm collection (Plant Genetic Resources Conservation Unit, Southern Regional Plant Introduction Station, Griffin, GA) (Table 1). Seed of the cultivar Crimson Sweet and the selection Egun were obtained from Johnny’s Selected Seed (Winslow, ME) and G.E. Boyhan (University of Georgia, Athens, GA), respectively. Near magnetic resonance [NMR (MiniSpec MQ20 NMR analyzer; Bruker Optics, Billerica, MA)] was used to determine the composition of egusi oil is comparable to safflower (Carthamus tinctorius), soybean (Glycine max), and sunflower (Helianthus annuus) oil as a feedstock for biodiesel production (Giwa et al., 2010; Jarret and Levy, 2012).

Received for publication 15 May 2012. Accepted for publication 2 July 2012. This research was funded by Monsanto.

Corresponding author. E-mail: cmcgre1@uga.edu.
SOP as described by Burke et al. (2005) and Wills et al. (2010) for sunflower seeds but using watermelon seed standards. Briefly, at least 20 seeds (1.83 to 5.43 g) of each cultivar/accession were transferred to a flat-bottomed sample tube and total seed oil was determined as a percentage of seed weight.

An F2 mapping population of 187 individuals was developed using ‘Yamato-cream’ Strain II (Strain II; PI 279461) from Japan and a wild Nigerian egusi type [Egusi (PI 560023)] (Fig. 1). The development of the Strain II × Egusi F2 genetic map is described elsewhere (Prothro, 2010; Sandlin et al., 2012). Briefly, the map includes 357 SNP markers on 14 LGs with an average distance of 4.2 cM between markers.

The F2 population and parents were direct-seeded in the field at the University of Georgia’s Plant Science Farm in Watkinsville in Summer 2007. Plants were grown according to University of Georgia Cooperative Extension Service recommendations. One mature fruit from 142 individuals was harvested and seeds from each fruit were cleaned and collected by hand. Seed was scored as egusi or normal based on visual inspection and allowed to dry before determining SOP as described previously. Because SOP is a proportion of weight, the data were arcsine square root transformed before QTL analysis (Sokal and Rohlf, 1995; Wills et al., 2010).

After confirming mendelian inheritance fitting a 3:1 ratio, the normal trait was coded as a dominant marker (d,b) and mapped onto the existing Strain II · Egusi SNP map using JoinMap 4.0 (Kyazma, Wageningen, The Netherlands). Analysis for the detection of M-QTL was performed using composite interval mapping [CIM (Zeng, 1993, 1994)] and multiple interval mapping [MIM (Kao and Zeng, 1997; Kao et al., 1999; Zeng et al., 1999)] in WinQTL Cartographer (WinQTL Cart) Version 2.5 (Wang et al., 2011). Because the F2 plants in the field were open-pollinated, the population type was designated as “RF3” (Wang et al., 2010). For CIM, permutation tests (1000 permutations, $\alpha = 0.05$) were used to determine the threshold values for each trait (Churchill and Doerge, 1994; Doerge and Churchill, 1996). CIM analysis was performed using the standard model (Model 6) with a walk speed of 1 cM and forward–backward stepwise regression to set the number of marker cofactors. The cofactors within 10 cM on either side of the QTL were excluded from the model. E-QTL was detected using MIM, and significance was determined as recommended by the authors using the information criteria IC(k) = –2[log(L) – kc(n)/2] and penalty function c(n) = log(n). All LGs and QTL were visualized using MapChart 2.2 (Voorrips, 2002).

**Results and Discussion**

Total SOP of accesses with egusi seed type ranged from 30.30% to 40.60%, whereas values for the samples with normal seed ranged from 20.14% to 26.55% (Table 1). These results are

| Accession/cultivar* | Seed type | SOP (%) |
|---------------------|-----------|---------|
| PI 635617 ('New Hampshire Midget') | Normal | 20.14 |
| PI 635609 ('Klondike Black Seeded') | Normal | 22.29 |
| PI 279461 (Strain II) | Normal | 25.20 |
| PI 244019* | Normal | 25.37 |
| ‘Crimson Sweet’ | Normal | 27.05 |
| PI 593359 | Normal | 30.30 |
| PI 559999 | Egusi | 30.85 |
| PI 559993 | Egusi | 30.91 |
| PI 559992 | Egusi | 31.82 |
| PI 560018 | Egusi | 33.00 |
| PI 560008 | Egusi | 33.08 |
| PI 560017 | Egusi | 33.18 |
| PI 560004 | Egusi | 33.24 |
| PI 560000 | Egusi | 33.69 |
| PI 560014 | Egusi | 34.20 |
| PI 560019 | Egusi | 34.39 |
| PI 560003 | Egusi | 34.77 |
| PI 559997 | Egusi | 34.98 |
| PI 560002 | Egusi | 35.78 |
| PI 559996 | Egusi | 36.03 |
| PI 560005 | Egusi | 36.29 |
| PI 560010 | Egusi | 36.60 |
| PI 560020 | Egusi | 36.66 |
| PI 560007 | Egusi | 36.73 |
| PI 560012 | Egusi | 36.88 |
| PI 560001 | Egusi | 37.01 |
| PI 560015 | Egusi | 37.06 |
| PI 559995 | Egusi | 37.12 |
| PI 560009 | Egusi | 37.21 |
| PI 560006 | Egusi | 37.37 |
| PI 560016 | Egusi | 38.19 |
| PI 560013 | Egusi | 39.02 |
| PI 560011 | Egusi | 39.22 |
| Egun* | Egusi | 40.08 |
| PI 560024 | Egusi | 40.47 |
| PI 560023 | Egusi | 40.60 |

*Samples are ordered from lowest to highest SOP. The parents used for the mapping population are underlined.

‘Crimson Sweet’ was obtained from Johnny’s Selected Seed (Winslow, ME), whereas Egun was supplied by G.E. Boyhan (University of Georgia, Athens, GA). All other cultivars and accessions were obtained from the U.S. Department of Agriculture–Agricultural Research Service, National Plant Germplasm System, Plant Genetic Resources Conservation Unit, Southern Regional Plant Introduction Station, Griffin, GA.

*Citrullus lanatus var. citroides.

*Selection from PI 595203 (Ling et al., 2009; Murphy and Dane, 2009).
similar to results obtained in a recent study by Jarret and Levy (2012). Earlier research has reported seed oil content as high as 50% to 53% (Bankole et al., 2005; El-Adawy and Taha, 2001) in *C. lanatus*. However, the latter studies involved oil extracted from the kernel only, which is expected to have higher oil percentage than whole seed with a positive correlation ($r = 0.82$) reported between the two traits in sunflower (Leon et al., 1995).

In the Strain II $\times$ Egusi F$_2$ population, 42 fruit had the egusi seed phenotype, whereas 100 had the normal seed phenotype. A chi-square goodness-to-fit test showed that the trait fit a 3:1 ratio ($\chi^2 = 1.59$, df = 1, $P > 0.05$), confirming the results of Gusmini et al. (2004) that the egusi seed phenotype is controlled by a single recessive gene. The eg locus mapped between marker NW0248325 and NW0250248 to position 57.7 cM on LG 2 of the Strain II $\times$ Egusi genetic map (Fig. 2).

The SOP in the F$_2$ population ranged from 17.8% to 37.8% (Fig. 3). With the exception of two samples, all normal seed had SOP below 28%, whereas all egusi seed had SOP above 28%. CIM and MIM yielded similar results and both identified four M-QTL for SOP in the F$_2$ population (MIM results in Table 2). Three QTLs were identified on LG 2 and one on LG 9B. The QTL identified at the 57.7 cM position on LG 2 overlaps with the mapped position of the eg locus (Fig. 4) and is responsible for $\approx 84\%$ of the $R^2$. This result is in line with the distribution of the SOP phenotype in the population (Fig. 3), which shows a clear delineation between samples with the egusi seed phenotype and normal seed phenotype.

Two other M-QTLs were identified on LG 2, one at 42.6 cM and another at 81.0 cM (Table 2; Fig. 4). Unfortunately, there are few markers in the area surrounding the eg locus on LG 2 and the distance between markers is large. This makes it difficult to determine whether these two M-QTLs truly represent additional M-QTL on LG 2 or whether they are artifacts of the major effect eg locus. In an effort to elucidate this matter, the F$_2$ population was divided based on seed type (normal or egusi) and mapped as two separate populations. This approach eliminated the effect of the eg locus, but also drastically lowered the size for mapping populations. In the normal seed population ($n = 100$), an M-QTL

---

Table 2. Main effect quantitative trait loci (M-QTL) and epistatic effect QTL (E-QTL) identified for seed oil as a percentage of seed weight (SOP) in the Strain II (PI 279461) $\times$ Egusi (PI 560023) F$_2$ watermelon population using multiple interval mapping (MIM) (Kao and Zeng, 1997; Kao et al., 1999; Zeng et al., 1999).

| M-QTL      | LG  | Position (cM) | LOD  | $R^2$ (%) | Additive effect | Dominance effect | LOD-1 support interval (cM) | LOD-1 support interval (cM) |
|------------|-----|---------------|------|-----------|----------------|------------------|---------------------------|---------------------------|
|            | 2   | 42.6          | 1.90 | -5.95     | 0.0091         | 0.0096           | 31.1                      | 51.0                      |
|            | 2   | 57.7          | 18.94| 83.88     | -0.0478        | -0.0724          | 54.9                      | 61.9                      |
|            | 2   | 81.0          | 4.50 | 9.26      | -0.0147        | -0.0025          | 78.7                      | 84.6                      |
|            | 9B  | 86.2          | 3.48 | 1.31      | 0.0059         | 0.0144           | 77.3                      | 95.1                      |

| E-QTL      | LG Type of interaction | LOD  | $R^2$ (%) | Phenotypic effect |
|------------|------------------------|------|-----------|------------------|
|            | 2 (57.7) $\times$ 9B D $\times$ A | 2.31 | 0.7       | -0.0198          |

*Linkage group.
*Log$_{10}$ likelihood ratio.
*Phenotypic variation explained ($R^2$).
*Negative values indicate that the effect is contributed by the allele from PI 560023 (Egusi). The results are for the arcsine square root transformed data.
*Dominant $\times$ additive.
QTLs associated with SOP have been described in cruciferous (Vollmann and Rajcan, 2010). Epistatic interactions between [2006; Zhao et al., 2005), soybean (Lark et al., 1994), and oats tease out potential QTL in addition to the watermelon. Future research will aim to add markers to LG 2 to and identified M-QTL and E-QTL associated with SOP in oilseed rape (Brassica sp.) oil crops (Delourme et al., 2006; Mahmood et al., 2006; Zhao et al., 2005), soybean (Lark et al., 1994), and oats [Avena sativa (Zhu et al., 2004)].

We have mapped the location of the eg locus in watermelon and identified M-QTL and E-QTL associated with SOP in watermelon. Future research will aim to add markers to LG 2 to tease out potential QTL in addition to the eg locus as well as make eventual cloning of the eg gene possible. Jarret and Levy (2012) recently reported a high correlation between the hull/kernel ratio and SOP in watermelon, showing that egusi seed has a significantly lower hull/kernel ratio than normal seed. We are currently pursuing this line of research as well as the potential contribution of the paternal genotype to SOP.

Literature Cited

Achigan-Dako, E.G., R. Fagbemissi, H.T. Avohou, R.S. Vodouhe, O. Coulibaly, and A. Ahanchede. 2008. Importance and practices of egusi crops [Citrullus lanatus (Thunb.) Matsum. & Nakai, Cucumeropsis manniit Naudin and Lagenaria siceraria (Molina) Standl. cv. ‘Aklamka’] in sociolinguistic areas in Benin. Biotechnol. Agron. Soc. Environ. 12:393–403.

Achu, M.B., E. Fokou, C. Tchiegang, M. Fotso, and F.M. Tchouanguep. 2005. Nutritive value of some cucurbitaceae oilseeds from different regions in Cameroon. Afr. J. Biotechnol. 11:1329–1334.

Bankole, S.A., A. Osho, A.O. Joda, and O.A. Enikuomehin. 2005. Effect of drying method on the quality and storability of ‘egusi’ melon seeds (Cucumis melo citrullus L.). Afr. J. Biotechnol. 4:799–803.

Burke, J.M., S.J. Knapp, and L.H. Rieseberg. 2005. Genetic consequences of selection during the evolution of cultivated sunflower. Genetics 171:1933–1940.

Che, K.-P., C.-Y. Liang, Y.-G. Wang, D.-M. Jin, B. Wang, Y. Xu, G.-B. Kang, and H.-Y. Zhang. 2003. Genetic assessment of watermelon germplasm using the AFLP technique. HortScience 38:81–84.

Churchill, G.A. and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963–971.

Dane, F. and P. Lang. 2004. Sequence variation at cpDNA regions of watermelons and related wild species: Implications for evolution of Citrullus haplotypes. Amer. J. Bot. 91:1922–1929.

Dane, F. and J. Liu. 2007. Diversity and origin of cultivated and citron type watermelon (Citrullus lanatus). Genet. Resources Crop Evol. 54:1255–1265.

Delourme, R., C. Falesnt, V. Huteau, V. Clouet, R. Horvais, B. Gandon, S. Specel, L. Hannef, J. Dheu, M. Deschamps, E. Margale, P. Vincourt, and M. Renard. 2006. Genetic control of oil content in oilseed rape (Brassica napus L.). Theor. Appl. Genet. 113:1331–1345.

Doerge, R.W. and G.A. Churchill. 1996. Permutation tests for multiple loci affecting a quantitative character. Genet. Mol. Biol. 142:285–294.
El-Adawy, T.A. and K.M. Taha. 2001. Characteristics and composition of watermelon, pumpkin, and paprika seed oils and flours. J. Agr. Food Chem. 49:1253–1259.

Fursa, T.B. 1972. K sistematic rodta Citrullus Schrad. Bot. Z. 57:31–41.

Giwa, S., L.C. Abdullah, and N.M. Adam. 2010. Investigating “egusi” (Citrullus colocynthis L.) seed oil as potential biodiesel feedstock. Energies 3:607–618.

Gusmini, G., T.C. Wehner, and R.L. Jarret. 2004. Inheritance of egusi seed type in watermelon. J. Hered. 95:268–270.

Jarret, R.L. and I.J. Levy. 2012. Oil and fatty acid contents in seed of Citrullus lanatus Schrad. J. Agr. Food Chem. 60:5199–5204.

Kao, C.-H. and Z.-B. Zeng. 1997. Simple sequence repeats in watermelon [Citrullus lanatus (Thunb.) Matsum. et Nakai]. Genome 40:433–441.

Jeffrey, C. 2001. Cucurbitaceae, p. 1510–1557. In: Hanelt, P., R. Gusmini, G., T.C. Wehner, and R.L. Jarret. 2004. Inheritance of egusi seed type in watermelon. J. Hered. 95:268–270.

Kao, C.-H. and Z.-B. Zeng. 2002. Modeling epistasis of quantitative trait loci using Cockerham’s model. Genetics 160:1243–1261.

Kao, C.-H., Z.-B. Zeng, and R.D. Teasdale. 1999. Multiple interval mapping for quantitative trait loci. Genetics 152:1203–1216.

Lark, K.G., J. Orf, and L.M. Mansur. 1994. Epistatic expression of quantitative trait loci (QTL) in soybean [Glycine max (L.) Merr.] determined by QTL association with RFLP alleles. Theor. Appl. Genet. 88:486–489.

Leon, A.J., M. Lee, G.K. Rufener, S.T. Berry, and R.P. Mowers. 1995. Use of RFLP markers for genetic linkage analysis of oil percentage in sunflower seed. Crop Sci. 35:558–564.

Ling, K.-S., K. Harris, J. Meyer, A. Levi, N. Guner, T. Wehner, A. Bendahanme, and M. Havey. 2009. Non-synonymous single nucleotide polymorphisms in the watermelon eIF4E gene are closely associated with resistance to zucchini yellow mosaic virus. Theor. Appl. Genet. 120:191–200.

Mahmood, T., M. Rahman, G. Stringam, F. Yeh, and A. Good. 2006. Identification of quantitative trait loci (QTL) for oil and protein contents and their relationships with other seed quality traits in Brassica juncea. Theor. Appl. Genet. 113:1211–1220.

Murphy, J.F. and F. Dane. 2009. Evaluation of ‘AU-Performance’ watermelon for its response to virus inoculation. HortTechnology 19:609–612.

National Research Council of the National Academies. 2006. Lost crops of Africa: Volume II: Vegetables. National Academies Press, Washington, DC.

Nimmakayala, P., Y.R. Tomason, J. Jeong, S.K. Ponniah, A. Karunathilake, A. Levi, R. Perumal, and U.K. Reddy. 2010. Genetic reticulation and interrelationships among Citrullus species as revealed by joint analysis of shared AFLPs and species-specific SSR alleles. Plant Genet. Resour.; Characterization Util. 8:16–25.

Ntui, V., R. Khan, D. Chin, I. Nakamura, and M. Mii. 2010. An efficient Agrobacterium tumefaciens-mediated genetic transformation of “egusi” melon (Colocynthis citrullus L.). Plant Cell Tiss. Org. Cult. 103:15–22.

Prothro, J.M. 2010. Genetic mapping of phenotypic and quantitative trait loci underlying horticulturally important traits in watermelon. MS thesis, Univ. of Georgia, Athens, GA.

Robinson, R.W. and D.S. Decker-Walters. 1997. Cucurbits. CAB International Publishing, Wallingford, UK.

Sandlin, K.C., J.M. Prothro, A.F. Heesacker, N. Khalilian, R. Okashah, W. Xiang, E. Bachlava, D. Caldwell, D. Seymour, V. White, E. Chan, G. Tollia, C. White, D. Safran, E. Graham, S.J. Knapp, and C.E. McGregor. 2012. Comparative mapping in watermelon [Citrullus lanatus (Thunb.) Matsum. et Nakai]. Theor. Appl. Genet. (in press).

Sokal, R.R. and F.J. Rohlf. 1995. Biometry: The principles and practice of statistics in biological research. Freeman, New York, NY.

Vollmann, J. and I. Rajcan. 2010. Oil crops, Handbook of plant breeding. Springer, New York, NY.

Voorrips, R.E. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. 93:77–78.

Wang, S., C.J. Basten, P. Gaffney, and Z.B. Zeng. 2010. Windows QTL Cartographer 2.5 user manual. 15 Apr. 2011. <http://statgen.ncsu.edu/qtlcart/WinQTLCart.pdf>.

Wang, S., C.J. Basten, and Z.B. Zeng. 2011. Windows QTL Cartographer 2.5. Dept. of Statistics, North Carolina State Univ., Raleigh, NC.

Wehner, T. 2008. Watermelon, p. 381–418. In: Prohens, J. and F. Nuez (eds.). Vegetables I: Asteraceae, brassicaceae, chenopodicaceae, and cucurbitaceae. Springer, New York, NY.

Wills, D.M., H. Abdel-Haleem, S.J. Knapp, and J.M. Burke. 2010. Genetic architecture of novel traits in the Hopi sunflower. J. Hered. 101:727–736.

Zeng, Z.B. 1993. Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA 90:10972–10976.

Zeng, Z.B. 1994. Precision mapping of quantitative trait loci. Genetics 136:1457–1468.

Zeng, Z.B., C.H. Kao, and C.J. Basten. 1999. Estimating the genetic architecture of quantitative traits. Genet. Res. 74:279–289.

Zhao, J., H.C. Becker, D. Zhang, Y. Zhang, and W. Ecke. 2005. Oil content of a European × Chinese rapeseed population: QTL with additive and epistatic effects and their genotype environment interactions. Crop Sci. 45:51–59.

Zhu, S., B.G. Rossnagel, and H.F. Kaeppler. 2004. Genetic analysis of quantitative trait loci for groat protein and oil content in oat. Crop Sci. 44:254–260.