Development of β-carotene, lysine, and tryptophan-rich maize (Zea mays) inbreds through marker-assisted gene pyramiding

Neelima Chandrasekharan1,9, Nagalakshmi Ramanathan2,5,9, Bharathi Pukalenthy3,9, Sarankumar Chandran3,4,9, Dhasarathan Manickam5, Karthikeyan Adhimoolam3, Ganesan Kalipatty Nalliappan4, Sudha Manickam3, Ravikesavan Rajasekaran7, Vellaikumar Sampathrajan5, Vignesh Muthusamy8, Firoz Hossain8, Hari Shankar Gupta8 & Senthil Natesan4

Maize (Zea mays L.) is the leading cereal crop and staple food in many parts of the world. This study aims to develop nutrient-rich maize genotypes by incorporating \textit{crtRB1} and \textit{o2} genes associated with increased β-carotene, lysine, and tryptophan levels. UMI1200 and UMI1230, high quality maize inbreds, are well-adapted to tropical and semi-arid regions in India. However, they are deficient in β-carotene, lysine, and tryptophan. We used the concurrent stepwise transfer of genes by marker-assisted backcross breeding (MABB) scheme to introgress \textit{crtRB1} and \textit{o2} genes. In each generation (from F1, BC1F1–BC3F1, and ICF1–ICF3), foreground and background selections were carried out using gene-linked (\textit{crtRB1} 3′TE and umc1066) and genome-wide simple sequence repeats (SSR) markers. Four independent BC2F1 lines of UMI1200 × CE477 (Cross-1), UMI1200 × VQL1 (Cross-2), UMI1230 × CE477 (Cross-3), and UMI1230 × VQL1 (Cross-4) having \textit{crtRB1} and \textit{o2} genes and 87.45–88.41% of recurrent parent genome recovery (RPGR) were intercrossed to generate the ICF1–ICF3 generations. Further, these gene pyramided lines were examined for agronomic performance and the β-carotene, lysine, and tryptophan contents. Six ICF3 lines (DBT-IC-β1σ4-4-8-8, DBT-IC-β1σ4-9-21-21, DBT-IC-β1σ4-10-1-1, DBT-IC-β2σ5-9-51-51, DBT-IC-β2σ5-9-52-52 and DBT-IC-β2σ5-9-53-53) possessing \textit{crtRB1} and \textit{o2} genes showed better agronomic performance (77.78–99.31% for DBT-IC-β1σ4 population and 85.71–99.51% for DBT-IC-β2σ5 population) like the recurrent parents and β-carotene (14.21–14.35 μg/g for DBT-IC-β1σ4 and 13.28–13.62 μg/g for DBT-IC-β2σ5), lysine (0.31–0.34% for DBT-IC-β1σ4 and 0.31–0.33% for DBT-IC-β2σ5), and tryptophan (0.079–0.085% for DBT-IC-β1σ4 and 0.078–0.083% for DBT-IC-β2σ5) levels on par with that of the donor parents. In the future, these improved lines could be developed as a cultivar for various agro-climatic zones and also as good genetic materials for maize nutritional breeding programs.

1Department of Plant Biotechnology, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore 641003, Tamil Nadu, India. 2Department of Plant Breeding and Genetics, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore 641003, Tamil Nadu, India. 3Department of Plant Breeding and Genetics, Agricultural College and Research Institute, Tamil Nadu Agricultural University, Madurai 625105, Tamil Nadu, India. 4Department of Plant Molecular Biology and Bioinformatics, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore 641003, Tamil Nadu, India. 5Department of Biotechnology, Agricultural College and Research Institute, Tamil Nadu Agricultural University, Madurai 625105, Tamil Nadu, India. 6Department of Millets, Center for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore 641003, Tamil Nadu, India. 7Department of Forage Crops, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore 641003, Tamil Nadu, India. 8Division of Genetics, Indian Agricultural Research Institute, New Delhi, India. 9These authors contributed equally: Neelima Chandrasekharan, Nagalakshmi Ramanathan, Bharathi Pukalenthy and Sarankumar Chandran. *email: senthil_natesan@tnau.ac.in
Maize, an important cereal, is life for millions in the global population, as a source of protein, vitamins, minerals, oils, and dietary fibre. The crop is cultivated widely in diverse agroecology across the globe and has the highest genetic yield potential among the cereals. It is grown in more than 160 countries with a total production of 1.05 million thousand tonnes and 28.90 million tonnes in India for the year 2019. Maize is a rich source of provitamin A and non-provitamin A carotenoids. The carotenoids are synthesized in the maize endosperm via the carotenoid biosynthesis pathway that originates from the isoprenoid precursor, geranyl pyrophosphate, supplied by the MEP pathway. Through a series of enzyme-mediated reactions, phytoene, the first carotenoid molecule, is synthesized and enzymatically converted to lycopene. This is the branch point of the pathway, and further conversion depends on the cyclization of the lycopene molecule. An asymmetric cyclization would produce an α-carotene molecule, and a symmetric cyclization would yield a β-carotene molecule, forming the carotenoid precursor for provitamin A. Several studies have shown that variations in the expression of genes directly involved in influencing the β-carotene levels in the maize endosperm are a pressing need for alleviating this micronutrient complication, and since the carotenoid compounds are naturally accumulated in the edible part of the maize endosperm, it becomes an ideal crop for biofortification.

Several studies have identified various genes that are directly involved in the variation of the β-carotene pathway by directly or indirectly modifying the carotenoid biosynthesis pathway. The LcyE and the crtRB1 genes were shown to be directly involved in influencing the beta carotene levels in the maize endosperm. The precise manipulation of the crtRB1 gene has shown to favorably increase the beta carotene concentration in previous studies. Yan et al. identified the crtRB1 gene responsible for this conversion and also three polymorphisms that influence the variation in the carotenoid concentration. The polymorphism in the 3’TE region with the favorable allele (543 bp) increases the carotene concentration in maize.

Maize also contains two protein fractions viz., zein and non-zein, where zein proteins are predominant. However, these zein proteins lack essential amino acids like lysine and tryptophan and hence induce Protein Energy Malnutrition (PEM). Several natural mutants (i.e., opaque 2 (o2), floury 2 (fl2), opaque 7 (o7), opaque 6 (o6), floury 3 (fl3)) have shown to increase these essential amino acids in maize of which o2 has been widely studied. The o2 mutant is known to decrease the zein fraction and increase the non-zein fraction which is naturally high in essential amino acids. The large genetic variation present in maize makes it an ideal crop for nutritional improvement specifically in regard to micronutrient deficiencies. Marker-assisted backcross breeding (MABB) has been shown to be a promising technique to introgress several nutritionally important genes in many crops including maize. Nutritional traits viz., provitamin A, higher protein content, high Zn, Fe, and Se content have been improved in maize through the MABB technique. Several studies in India and other parts of the world have successfully introgressed either crtRB1 or o2 into popular elite lines and improved the β-carotene, lysine, and tryptophan contents. The common determinant in all the previous studies is the introgression of a single factor into an established variety. By adopting

Abbreviations

| Abbreviation | Description |
|--------------|-------------|
| EL | Ear length |
| EW | Ear weight |
| DS | Days to silking |
| DT | Days to tasseling |
| EH | Ear height |
| HPLC | High performance liquid chromatography |
| KW | Kernel weight |
| LB | Leaf breadth |
| LL | Leaf length |
| MABB | Marker assisted backcross breeding |
| MAGP | Marker assisted gene pyramiding |
| NKR | Number of kernel rows |
| NKRE | Number of kernel rows per ear |
| NTL | Number of tassel branches |
| PH | Plant height |
| PCR | Polymerase chain reaction |
| PME | Protein energy malnutrition |
| RPGR | Recurrent parent genome recovery |
| SPY | Single plant yield |
| SSR | Simple sequence repeat |
| TL | Tassel length |
| UNICEF | United Nations International Children’s Emergency Fund |
| USA | United States of America |
| VAD | Vitamin A deficiency |
the technique of gene pyramiding, varieties can be produced with broad sense capabilities and essentially more important genetic stocks. Especially by bringing improved versions of β-carotene, lysine, and tryptophan into a single genotype, the time required to improve the plants individually is reduced and would also provide a superior genotype with several favourable nutritional traits. This has now become possible due to the advances made in technology as well as the identification of new molecular markers and integrated techniques developed for efficient selection. Considering these, this study is planned to develop an intercross population and pyramid the crtRB1 and o2 simultaneously in the background of elite genotypes.

Results

Transfer of crtRB1 and o2 genes into UMI1200. A total of 27 and 23 F1s were produced in cross-1 and cross-2, and their heterozygosity was confirmed via foreground markers associated with crtRB1 and o2 genes. The healthy F1s from both crosses were backcrossed with a recurrent parent to produce 106 and 232 BC2F1 lines, and all heterozygous conditions were confirmed in BC2F1 lines using foreground markers. All the heterozygous lines were subjected to background selection with 112 and 106 polymorphic markers. They showed 52.82–56.41% and 62.13–74.25% of RPGR with an average of 54.84% and 69.38% in cross-1 and cross-2. Further, one BC2F1 line from each cross with crtRB1 and o2 genes and maximum RPGR was selected and backcrossed with a recurrent parent to produce 136 and 218 BC2F1 lines. Following similar selection procedures, BC2F1 generation was advanced to BC2F1. A total of 85 and 109 BC2F1 lines were produced to the cross-1 and cross-2, and foreground selection revealed that the 24 and 31 BC2F1 lines had crtRB1 and o2 genes in the heterozygous condition. All these lines were subjected to background selection, and they exhibited 86.35–88.52% and 86.14–88.21% of RPGR with an average of 87.74 and 87.45% (Supplementary Tables S1, S2). Among them, two lines, DBT 1-1-1-17-5-14 from cross-1 and DBT 4-1-1-10-10-16 from cross-2 having maximum RPGR, were used to develop the intercross population (designated as DBT-IC-β1σ4, o2).

Transfer of crtRB1 and o2 genes into UMI1230. With the support of foreground markers, crtRB1 and o2 genes heterozygous lines were confirmed in F1s from cross-3 and cross-4. The F1s were backcrossed with a recurrent parent to produce 121 and 160 BC3F1 lines. Among them, 42 and 68 BC3F1 lines possessing crtRB1 and o2 genes in their heterozygous condition were identified in cross-3 and cross-4 using foreground markers and were subjected to background selection with 114 and 90 polymorphic SSR markers. Background selection revealed 53.87–57.69% and 68.60–76.20% of RPGR with an average of 55.12% and 72.70%. One BC3F1 line from each cross possessing crtRB1 and o2 genes and maximum RPGR was selected and backcrossed with a recurrent parent to produce 146 and 153 BC3F1 lines. Applying the same strategy, 5 BC3F1 and 10 BC3F1 lines possessing crtRB1 and o2 genes and maximum RPGR were identified. The BC3F1 lines from cross-3 and cross-4 exhibited 86.75–88.84% and 87.56–89.42% of RPGR with an average of 87.84% and 88.41% (Supplementary Tables S1, S2). The two BC3F1 lines, (DBT 2-1-4-7-1-9) and (DBT 5-1-14-5-8-7) from cross-3 and 4 having maximum RPGR, were used to develop the intercross population (designated as DBT-IC-β2σ5, o2).

Stacking of crtRB1 and o2 genes. The line DBT 1-1-1-17-5-14 (derived from cross 1) was used as the female parent and DBT 4-1-1-10-10-16 (derived from cross 2) as the male parent in the development of intercross population (DBT-IC-β1σ4, o2) to pyramid crtRB1 and o2 genes. Among the 128 ICF1 lines, 64 lines were confirmed to be heterozygous for two target genes. Of these, 64 ICF1 were selected and selfed to obtain 40 ICF2 lines. Foreground selection was conducted in ICF2 lines to trace the lines carrying a combination of two genes. Based on foreground selection and the phenotyping of kernels for opaqueness (25%), a total of 9 homozygous lines with crtRB1 and o2 genes were identified. Chi-square test on the 9 lines revealed that the population followed the expected Mendelian ratio of 1:2:1 (Table 1; Fig. 1). Background selection was done in those selected 9 lines with 148 polymorphic SSR markers and selfed to produce ICF3 generation (Supplementary Table S3). Eventually, 3 lines, DBT 1-1-1-17-5-14 from cross 1 and DBT 4-1-1-10-10-16 from cross 2 having maximum RPGR, were used to develop the intercross population (designated as DBT-IC-β1σ4).

Evaluation of ICF2 generation for morphological traits. The newly developed 6 ICF2 line's agronomical performance was evaluated (Fig. 1) by measuring 14 morphological traits and estimating the similarity percentage compared to the recurrent parent (Tables 2, 3). The three improved lines DBT-IC-β1σ4-4-8-8, DBT-IC-β1σ4-9-21-21, and DBT-IC-β1σ4-10-1-1 from the DBT-IC-β1σ4 population showed more than 90% similarity to the recurrent parent UMI1200 for most of the traits. The same was the case for the three improved lines from DBT-IC-β2σ5, o2. In the DBT-IC-β1σ4 population, the similarity percentage ranged from 77.78% (NTB) to 99.31% (LB), and DBT-IC-β1σ4-4-8-8 showed the highest similarity percentage of 99.31% for LB, followed by DBT-IC-β1σ4-9-21-21 showing a similarity percentage of 99.28% for EW. In the DBT-IC-β2σ5, o2 population, the similarity percentage ranged from 85.71% (NTB and NKRE) to 99.51% (LL), and DBT-IC-β2σ5, o2-9-53-53 had the highest similarity percentage of 99.51% for LB.
| a2           | Genotypic class | χ² | P-value | Genotypic class | χ² | P-value |
|-------------|----------------|----|---------|----------------|----|---------|
| DBT-IC-β1σ4 | O2O2           | 0.400 ns | 0.819 | Allele3        | 0.618 ns | 0.734 |
| DBT-IC-β1σ4 | O2o2           | 0.432 ns | 0.806 | Allele3/Allele1 | 0.789 ns | 0.674 |
| DBT-IC-β1σ4 | o2o2           | 0.525 ns | 0.772 | Allele1        | 0.687 ns | 0.647 |
| DBT-IC-β2σ5 | O2O2           | 0.400 ns | 0.819 | Allele3        | 0.618 ns | 0.734 |
| DBT-IC-β2σ5 | O2o2           | 0.432 ns | 0.806 | Allele3/Allele1 | 0.789 ns | 0.674 |
| DBT-IC-β2σ5 | o2o2           | 0.525 ns | 0.772 | Allele1        | 0.687 ns | 0.647 |

**Table 1.** Segregation pattern of o2 and crtRB1 allele in intercross (IC) population of DBT-IC-β1σ4 and DBT-IC-β2σ5. O2O2, Homozygous dominant; O2o2, Heterozygotes; o2o2, Homozygous recessive (Favourable); Allele 3 (Unfavourable); Allele 3/1 (Unfavourable); Allele 1 (Favourable); ns (Non significant).

Similarity percentage of 99.51% (LL), followed by DBT-IC-β1σ4-9-52-52 having a similarity percent of 98.86% for EL.

β-carotene, lysine, and tryptophan contents in ICF3 lines. The β-carotene content in the recurrent parents, UM11200 and UM11230, was found to be 0.60 µg/g and 1.20 µg/g respectively, whereas, for the donor parent, CE477, the β-carotene content was found to be 15.20 µg/g. In the DBT-IC-β1σ4 population, the highest β-carotene content was found in DBT-IC-β1σ4-9-21 (10.35 µg/g) followed by DBT-IC-β1σ4-10-1 (10.25 µg/g) and DBT-IC-β1σ4-9-5 (10.20 µg/g). The β-carotene content was found to be significantly higher in the DBT-IC-β1σ4 population compared to the recurrent parent VQL1. In the DBT-IC-β2σ5 population, the highest β-carotene content was found to be 10.25 µg/g in DBT-IC-β2σ5-9-21-21, which is comparable with that of the donor parent. This was followed by DBT-IC-β2σ5-9-52-52 having a similarity percent of 98.86% (Table 4).

Discussion
To improve maize lines with β-carotene, lysine, and tryptophan, in our study, we were able to pyramid two nutritionally important genes crtRB1 and o2 into a single genotype by way of intercrossing. In our breeding programme, four independent crosses (UM11200 × CE477, UM11200 × VQL1, UM11230 × CE477, and UM11230 × VQL1) were formed to incorporate the crtRB1 and o2 genes into two elite inbred lines UM11200 and UM11230. For the marker assisted backcrossing, selection using crtRB1 gene-specific and umc1066 markers was done to identify four BC2F1 lines from each cross having the desired genes and maximum RPG%. The lines DBT-1-1-17-5-14 from cross-1 and DBT-4-1-10-10-16 from cross-2 were intercrossed to produce the DBT-IC-β1σ4 population, which in hindsight improved UM11200 for β-carotene, lysine, and tryptophan levels. Similarly, the lines DBT-2-1-4-7-1-9 from cross-3 and DBT-5-1-14-5-8-7 from cross-4 were intercrossed to produce the DBT-IC-β2σ5 population, which improved UM11230 for β-carotene lysine and tryptophan levels.

In all the IC generations (ICF1–ICF3), the same markers were used to ensure that the final products were double homozygotes for both the crtRB1 and o2 genes. In the ICF1 generation, generated lines from both DBT-IC-β1σ4 and DBT-IC-β2σ5 populations were subjected to the chi-square test. The results revealed that the population segregated in the expected Mendelian ratio of 1:2:1 without any significant distortion for both the markers.

https://doi.org/10.1038/s41598-022-11585-y

Scientific Reports | (2022) 12:8551 | https://doi.org/10.1038/s41598-022-11585-y

natureportfolio
Similar results were also obtained by Veldboom and Lee,35 and Lu et al.36 The selected double positive lines were then used to produce the ICF$_1$ generation wherein the double homozygotes were ensured using the $crtRB1$ 3′TE gene-specific and umc1066 markers. In this way, we were able to stack the nutritionally important genes and develop lines that were improved for β-carotene, lysine, and tryptophan levels. Similar studies were reported by other researchers.24,26,31,33,37 However, in our study, we were able to achieve gene stacking by intercrossing homogenous lines that already had enhanced levels of β-carotene, lysine, and tryptophan thereby reducing the breeding cycle due to which we were able to produce a homogenous population that was highly similar to that of the recurrent parent in a short amount of time. Moreover, we were able to improve UMI1200 and UMI1230 which are the parents of a popular maize hybrid CO6 that is most suited to the climatic regions of South India.

Recovery of recurrent parent genome was also achieved in both ICF$_2$ and ICF$_3$ generation using a total of 148 polymorphic SSR markers. A high RPG% was obtained in the ICF$_2$ generation itself due to the initial improved lines used to produce the intercross population having low levels of unwanted linkage drag. Once the ICF$_3$ generation was developed we were able to identify three lines in both cross combinations that were double homozygotes and had a high recovery of recurrent parent genome. These results are in accordance with earlier reports19,26,32. The analysis of the opaqueness in the ICF$_2$ generation showed that all the seeds showed 25% opaqueness for both the cross combinations. This was achieved because the lines that were used to produce the intercross population were already established for the 25% opaqueness using the lightbox screening method. Therefore, the progenies of the ICF$_3$ generation also showed only 25% opaqueness. These results are in accordance with the previous findings24,26,31,33,34,37.

Morphological trait evaluation in the ICF$_1$ generation for both DBT-IC-β$_1$σ$_4$ and DBT-IC-β$_2$σ$_5$ populations revealed that the improved lines were having more than 90% similarity with that of the recurrent parent without any major differences in important yield characters like SPY and EW. It showed that complete recovery of important phenotypic and yield characters of the recurrent parent was attained in the pyramided lines along with the desired genes. The lines DBT-IC-β$_1$σ$_4$-10-1-1 and DBT-IC-β$_1$σ$_4$-9-21-21 from the DBT-IC-β$_1$σ$_4$ population and the lines DBT-IC-β$_2$σ$_5$-9-51-51 and DBT-IC-β$_2$σ$_5$-9-52-52 from the DBT-IC-β$_2$σ$_5$ population were found to have the highest similarity to the respective recurrent parents as far as the yield characters were concerned. Similar results were also reported by former researchers24,34,40.

The evaluation of nutritional contents proved that the ICF$_1$ lines had improved levels of β-carotene, lysine, and tryptophan levels in comparison with their normal recurrent parents. In the DBT-IC-β$_1$σ$_4$ population, DBT-IC-β$_1$σ$_4$-9-21-21 and DBT-IC-β$_1$σ$_4$-4-8-8 had the highest levels of β-carotene, lysine, and tryptophan respectively. Whereas, in the DBT-IC-β$_2$σ$_5$ population, DBT-IC-β$_2$σ$_5$-9-51-51 and DBT-IC-β$_2$σ$_5$-9-53-53 had the highest levels of β-carotene, lysine, and tryptophan respectively. Similar results were also obtained by earlier studies26,28,33,37. The improved lines in both cross combinations obtained from the ICF$_2$ generation not only have the target donor genes with elevated nutrition levels but also has the high recovery of recurrent parent genome as well as highest phenotypic similarity to that of the recurrent parents rendering them crucial genetic materials for further hybrid synthesis and other genetic studies.

The present study has resulted in the development of improved lines possessing two genes ($crtRB1$ and o2) responsible for β-carotene, lysine, and tryptophan by marker-assisted gene pyramiding (MAGP) strategy. Thus, the pyramided inbred lines (UMI 1200 and UMI 1230) recorded a higher level of β-carotene, lysine, and tryptophan thereby reducing the breeding cycle due to which we were able to produce a homogenous population that was highly similar to that of the recurrent parent in a short amount of time. Moreover, we were able to improve UMI1200 and UMI1230 which are the parents of a popular maize hybrid CO6 that is most suited to the climatic regions of South India.

Materials and methods

Plant genetic materials. Maize inbreds, UMI1200, and UMI1230, well-adapted to tropical and semi-arid regions in India were selected as the recurrent parents. Because of their good combining ability, both were utilized to develop the CO6 hybrid. The inbreds seeds were obtained from the Department of Plant Genetic Resources, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore. VQL1 (Posessing o2 associated with high lysine and tryptophan contents) and CE477 (Posessing $crtRB1$ associated with high β-carotene content) were selected as donor parents. VQL1 was obtained from Vivekananda Parvatiya Krishi Anusandhan Sansthan (VPKAS), Almora, India, whereas CE477 was obtained from International Maize and Wheat Improvement Center, Mexico.

Foreground and background selection. Foreground selection was done using closely linked markers to $crtRB1$ and o2 genes. The $crtRB1$ gene located in chromosome 10 was selected using InDel marker $crtRB1$ 3′TE, whereas the o2 gene located in chromosome 7 was selected using the simple sequence repeat (SSR) marker umc106641. The background selection was done to examine the recurrent parent genome recovery (RPGA). It was performed using 248 SSR markers with known chromosomal positions distributing all ten maize chromosomes. All primer sequences were obtained from the maize genome database (www.maizegdb.org) and synthesized by Eurofins Ltd., Bangalore, India.

DNA extraction and PCR amplification. Genomic DNA was isolated from a two-week-old plant following the method by Murray and Thompson42. The PCR analysis for the $crtRB1$ gene-specific marker $crtRB1$ 3′TE (65F: ACACACATGGACAAGTCTCG and 62R: ACACACATGGCGATCAGGACC) was carried out in a 10 μl reaction containing 2 μl of 20 ng template DNA, 2 mM of MgCl$_2$, 1 mM of dNTPs, 2 μM of primer pair and 1.5U of Taq polymerase. The screening followed the ‘touch down’ technique of an initial denaturation for 5 min at 94 °C, followed by 19 cycles of denaturation for 45 s at 94 °C, annealing for 30 s at 62 °C with a reduction of 0.5 °C in every cycle down to 54 °C and extension for 1 min at 72 °C followed by
another 20 cycles of denaturation at 94 °C for 45 s, annealing at 54 °C for 30 s, extension at 72 °C for 1 min and 20 s and a final extension at 72 °C for 10 min. The PCR analyses for the \( \alpha_2 \) gene-specific marker umc1066 (62R: ACACCTCGGACCAGAACC, 66R: AAGCCAATACAGGGGACCA) and other background SSR markers were carried out in a 10 μl reaction containing 2 μl of 20 ng template DNA, 2 mM of MgCl₂, 1 mM of dNTPs, 2 μM of primer pair, and 1.5U of Taq polymerase. The template DNA underwent an initial denaturation at 94 °C for 7 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 60 °C for 30 s and extension at 72 °C for 45 s followed by a final extension at 72 °C for 7 min. The amplified PCR products were run using a 3% agarose gel for 3 h with the addition of 5 μl bromophenol blue, and the resolution was documented after 3 h.

**Marker aided transfer of \( \alpha rB1 \) and \( \alpha 2 \) genes.** Four crossing programs, UM11200 × CE477 (Cross-1), UM11200 × VQL1 (Cross-2), UM11230 × CE477 (Cross-3), and UM11230 × VQL1 (Cross-4) were initiated to develop the nutrients rich lines using UM11200 and UM11230 (Recipient) and CE477 and VQL1 (Donor) (Fig. 2). The \( F_5 \) from all the four crosses were verified for the existence of \( \alpha rB1 \) and \( \alpha 2 \) genes in heterozygous form with foreground markers and then backcrossed with UM11200 or UM11230 to produce \( BC_{GF} \). The \( BC_{GF} \) lines having \( \alpha rB1 \) (Cross-1and 3) and \( \alpha 2 \) (Cross-2 and 4) in heterozygous form were selected with foreground markers. The foreground positives from \( BC_{GF} \) were subjected to background selection to identify the plants with maximum recovery of recurrent parent genome using polymorphic SSR markers. Similarly, another two rounds of backcrossing followed by foreground and background selection generated \( BC_{GF} \) lines having \( \alpha rB1 \) (Cross-1 and 3) and \( \alpha 2 \) (Cross-2 and 4) with maximum recovery of recurrent parent genome. The final lines were crossed to produce intercross \( F_5 \) (ICF₁) to combine the \( \alpha rB1 \) and \( \alpha 2 \) genes into a single plant. The heterozygous form in ICF₁ was confirmed by foreground markers and then selfed to two generations to produce ICF₂. The ICF₂ and ICF₃ generations were subjected to the foreground and background selection.

**Observation of kernel modification via lightbox screening.** The \( \alpha o2 \) allele that is associated with the increased lysine and tryptophan content is also associated with an undesirable character of kernel softness that can be visualized as opaqueness in the kernels. Based on the opaqueness, the kernels can be categorized into five levels: 0%, 25%, 50%, 75%, and 100%. Usually, 25% and 50% kernels are selected since they are certain to contain the \( \alpha o2 \) gene in a homozygous recessive state. Whereas, the other categories contain the \( \alpha 2 \) gene in either heterozygous or homozygous dominant condition and are heavily susceptible to unfavourable irregularities. A lightbox apparatus is used to differentiate the level of kernel opaqueness as an indirect measure of the kernel softness. Hence, by the dual selection technique of lightbox screening and foreground selection, the \( \alpha o2 \) allele is guaranteed in the population. The ICF₂ and ICF₃ generation lines were subjected to the lightbox screening and the lines exhibiting 25% opaqueness are selected to fix the \( \alpha 2 \) allele in the homozygous recessive state.

**Characterization of ICF₃ lines for morphological traits.** The newly developed intercross lines from both the cross combinations were planted along with the donor and recurrent parent. The plants were maintained with a distance of 20 cm, row spacing of 60 cm, and a row length of 3 m. Good agronomic practices were maintained during the growing period of the crop. Randomized Block Design (RBD) was performed with three replication. Randomly five plants were selected for the morphological trait evaluation. The recovery percentage of the recurrent parents was calculated according to the previous researchers³⁹, ³³. The plants were examined for the agronomic performance by measuring 14 morphological characters viz., days to tasseling (DT, in days), days to silking (DS, in days), plant height (PH, cm), ear height (EH, cm), tassel length (TL, cm), number of tassel branches (NTB), leaf length (LL, cm), leaf breadth (LB, cm), ear length (EL, cm), number of kernels rows per ear (NKRE), number of kernels per row (NKR), ear weight (EW, g), 100 kernel weight (KW, g) and single plant yield (SPY, g). All the characterizations were done according to the descriptors suggested by the International Board for Plant Genetic Resources⁴³.

**Analysis of \( \beta \) carotene, lysine, and tryptophan.** The kernels from the ICF₁ generation were examined for \( \beta \)-carotene, lysine, and tryptophan. The extraction of \( \beta \)-carotene was done following the method given by Kurilich and Juvik⁴⁴ and measured with the help of High-Performance Liquid Chromatography (HPLC). The final samples were eluted in a C30 column using a mobile phase consisting of acetonitrile: dichloromethane: methanol in the ratio of 75:20:5, and the flow rate was found to be 0.4 ml/min. The standard curve was constructed based on three different dilutions (1, 10, and 100 ppm) of standard beta carotene obtained from M/s Sigma Aldrich, USA. The lysine and tryptophan contents were measured following the colorimetric method⁴⁵.

---

(Fig. 1). Foreground and background selection and morphological traits evaluation in ICF₂ and ICF₃ populations. (a) Foreground selection of ICF₂ lines from DBT-IC-\( \beta \)-\( \alpha o \), using \( \alpha rB1 \) gene specific marker \( \alpha rB1 \) 3'TE, (M) Ladder (100 bp), (P₁) CE477, (P₂) UM11200, (1–10) ICF₂ plants; (b) Foreground selection of ICF₂ lines from DBT-IC-\( \beta \)-\( \alpha o \), using \( \alpha 2 \) gene linked marker umc1066, (M) Ladder (100 bp), (P₁) UM11200, (P₂) VQL1, (1–10) ICF₂ plants; (c) Foreground selection of ICF₃ lines from DBT-IC-\( \beta \)-\( \alpha o \), using \( \alpha 2 \) gene linked marker umc1066, (M) Ladder (100 bp), (P₁) UM11200, (P₂) CE477, (P₃) VQL1, (1–10) ICF₃ plants; (d) Foreground selection of ICF₃ lines from DBT-IC-\( \beta \)-\( \alpha o \), using \( \alpha 2 \) gene linked marker umc1066, (M) Ladder (100 bp), (P₁) UM11200, (P₂) CE477, (P₃) VQL1, (1–10) ICF₃ plants; (e–i) and (j), Background selection of ICF lines from DBT-IC-\( \beta \)-\( \alpha o \), (M) Ladder (100 bp), (P₁) UM11200, (P₂) CE477, (P₃) VQL1, (1–3) ICF₃ plants; (g, h, k and l), Background selection of ICF lines from DBT-IC-\( \beta \)-\( \alpha o \), (M) Ladder (100 bp), (P₁) UM11200, (P₂) CE477, (P₃) VQL1, (1–3) ICF₃ plants; (m–p), Evaluation of morphological traits in ICF₃ lines. UM1200 (m), UM11230 (n), DBT-IC-\( \beta \)-\( \alpha o \)-4-4-8-8 (o), DBT-IC-\( \beta \)-\( \alpha o \)-9-53-53 (p).
The samples were measured using the spectrophotometer at a wavelength of 390 nm for lysine and 560 nm for tryptophan, and the levels were expressed in percent\(^46\).

**Statement for the use of plant materials.** The study complies with local and national regulations.

**Table 2.** Comparison of the double positive lines in the ICF\(_3\) generation of DBT-IC-\(\beta_1\sigma_4\) along with its recurrent parents for the recovery percentage of morphological traits.

| Morphological traits | Recurrent parent | Identified positive lines | Recovery percentage (%) |
|----------------------|------------------|---------------------------|-------------------------|
|                      | DBT-IC-\(\beta_1\sigma_4\) | DBT-IC-\(\beta_1\sigma_4\) \(9 \rightarrow 21\) | DBT-IC-\(\beta_1\sigma_4\) \(10 \rightarrow 1\) | DBT-IC-\(\beta_1\sigma_4\) \(9 \rightarrow 21\) | DBT-IC-\(\beta_1\sigma_4\) \(10 \rightarrow 1\) |
| Days to tasseling (days) | UMI1200 | 58.00 | 56.00 | 58.00 | 55.00 | 57.00 | 96.55 | 94.83 | 98.28 |
| Days to silking (days) | 60.00 | 59.00 | 58.00 | 60.00 | 98.33 | 96.77 | 98.33 |
| Plant height (cm) | 155.87 | 154.68 | 152.89 | 153.00 | 99.24 | 98.09 | 98.16 |
| Ear height (cm) | 76.84 | 73.85 | 67.75 | 72.02 | 96.11 | 88.17 | 93.73 |
| Tassel length (cm) | 21.34 | 20.42 | 21.08 | 19.68 | 95.69 | 98.78 | 92.22 |
| Number of tassel branches | 9.00 | 7.00 | 7.00 | 8.00 | 77.78 | 77.78 | 88.89 |
| Leaf length (cm) | 57.24 | 56.66 | 55.31 | 54.53 | 98.99 | 96.63 | 95.27 |
| Leaf breadth (cm) | 7.20 | 7.15 | 6.87 | 7.06 | 99.31 | 95.42 | 98.06 |
| Ear length (cm) | 15.60 | 14.40 | 14.80 | 14.50 | 92.31 | 94.87 | 92.95 |
| Number of kernel rows per ear | 12.60 | 10.00 | 10.00 | 10.00 | 83.33 | 83.33 | 83.33 |
| Number of kernels per row | 26.00 | 23.00 | 24.00 | 24.00 | 88.46 | 92.31 | 92.31 |
| Ear weight (g) | 121.47 | 120.20 | 119.90 | 120.60 | 98.95 | 97.81 | 99.28 |
| 100 kernel weight (g) | 28.76 | 26.50 | 27.20 | 25.80 | 92.14 | 94.58 | 89.71 |
| Single plant yield | 99.98 | 96.71 | 98.31 | 95.86 | 96.73 | 98.33 | 95.88 |

**Figure 2.** Marker assisted backcrossing scheme (MABC) used to generate the intercross (IC) population. Cross 1 (UMI1200 × CE477); Cross 2 (UMI1200 × VQL 1); Cross 3 (UMI1230 × CE477); Cross 4 (UMI1230 × VQL 1). Crossing between parents (Kharif season, June to September 2015), F\(_1\) (Rabi season, November to March 2015–2016), BC\(_1\)F\(_1\) (Kharif season, June to September 2016), BC\(_2\)F\(_1\) (Rabi season, November to March 2016–2017), BC\(_3\)F\(_1\) (Kharif season, June to September 2017), ICF\(_1\) (Kharif season, June to September 2018), ICF\(_2\) (Rabi season, November to March 2018–2019), and ICF\(_3\) (Kharif season, June to September 2019).
Table 3. Comparison of the double positive lines in the ICF$_3$ generation of DBT-IC-$\beta_2\sigma_5$ along with its recurrent parents for the recovery percentage of morphological traits.

| Morphological traits | Recurrent parent | Identified positive lines | Recovery percentage (%) |
|----------------------|------------------|----------------------------|-------------------------|
|                      |                  | DBT-IC-$\beta_2\sigma_5$, 9-51-51 | DBT-IC-$\beta_2\sigma_5$, 9-52-52 | DBT-IC-$\beta_2\sigma_5$, 9-53-53 |
| Days to tasseling (days) | 60.00 | 57.00 | 58.00 | 59.00 | 95.00 | 96.67 | 98.33 |
| Days to silking (days) | 62.00 | 59.00 | 60.00 | 61.00 | 95.16 | 96.77 | 98.39 |
| Plant height (cm) | 158.40 | 156.28 | 153.94 | 150.25 | 98.66 | 97.18 | 94.85 |
| Ear height (cm) | 81.00 | 79.06 | 77.75 | 76.57 | 97.60 | 95.99 | 94.53 |
| Tassel length (cm) | 31.30 | 29.92 | 30.70 | 29.18 | 95.59 | 98.08 | 93.23 |
| Number of tassel branches | 14.00 | 12.00 | 13.00 | 12.00 | 85.71 | 92.86 | 85.71 |
| Leaf length (cm) | 63.20 | 58.19 | 62.11 | 62.89 | 92.07 | 98.28 | 99.51 |
| Leaf breadth (cm) | 7.50 | 7.20 | 7.10 | 7.20 | 96.00 | 94.67 | 96.00 |
| Ear length (cm) | 17.50 | 17.20 | 17.30 | 17.00 | 98.29 | 98.86 | 97.14 |
| Number of kernel rows per ear | 14.00 | 12.00 | 12.00 | 12.00 | 85.71 | 85.71 | 85.71 |
| Number of kernels per row | 25.00 | 24.00 | 22.00 | 24.00 | 96.00 | 88.00 | 96.00 |
| Ear weight (g) | 106.90 | 104.56 | 103.70 | 102.20 | 97.81 | 97.01 | 95.60 |
| 100 kernel weight (g) | 25.20 | 22.12 | 23.21 | 24.11 | 87.78 | 92.10 | 95.67 |
| Single plant yield | 75.51 | 72.56 | 73.74 | 71.11 | 96.89 | 97.66 | 94.17 |

Table 4. Lysine, tryptophan, and $\beta$-carotene levels of the ICF$_3$ improved double positive lines of DBT-IC-$\beta_1\sigma_4$ and DBT-IC-$\beta_2\sigma_5$.

| Trait | UMI1200 | UMI1230 | CE477 | VQL | DBT-IC-$\beta_1\sigma_4$ | DBT-IC-$\beta_2\sigma_5$ |
|-------|---------|---------|-------|-----|-------------------------|-------------------------|
|       |         |         |       |     | DBT-IC-$\beta_1\sigma_4$, 4-8-8 | DBT-IC-$\beta_1\sigma_4$, 9-21-21 | DBT-IC-$\beta_1\sigma_4$, 10-1-1 | DBT-IC-$\beta_2\sigma_5$, 9-51-51 | DBT-IC-$\beta_2\sigma_5$, 9-52-52 | DBT-IC-$\beta_2\sigma_5$, 9-53-53 |
| $\beta$-Carotene | 0.60 | 1.20 | 15.20 | 0.7 | 14.21 | 14.35 | 14.29 | 13.54 | 13.28 | 13.62 |
| Lysine | 0.26 | 0.25 | 0.13 | 0.42 | 0.33 | 0.31 | 0.32 | 0.34 | 0.32 | 0.31 |
| Tryptophan | 0.013 | 0.020 | 0.021 | 0.087 | 0.082 | 0.079 | 0.081 | 0.083 | 0.081 | 0.078 |

References
1. Knoema. "Maize Production Quantity". (New Knoema). https://knoema.com/atlas/topics/Agriculture/Crops-Production-Quantity-tonnes/Maize-production (2019).
2. Rodriguez-Concepcion, M. Supply of precursors for carotenoid biosynthesis in plants. Arch. Biochem. Biophys. 504, 118–122 (2010).
3. Cunningham, F. X. & Gantt, E. One ring or two? Determination of ring number in carotenoids by lycopene $\varepsilon$-cyclases. Proc. Natl. Acad. Sci. 98, 2905–2910 (2001).
4. North, H. M. et al. The Arabidopsis ABA-deficient mutant aba4 demonstrates that the major route for stress-induced ABA accumulation is via neoxanthin isomers. Plant J. 50, 810–824 (2007).
5. Sommer, A. & Davidson, F. R. Assessment and control of vitamin A deficiency: The Annecy Accords. J. Nutr. 132, 2845–2850 (2002).
6. WHO. Global database on vitamin A deficiency. Global prevalence of vitamin A deficiency in populations at risk. 1995–2005. http://whqlibdoc.who.int/publications/2009/9789241598019_eng.pdf (2009).
7. United Nations Children’s Fund. Vitamin A deficiency. https://data.unicef.org/topic/nutrition/vitamin-a-deficiency/ (2019).
8. Harjes, C. E. et al. Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319, 330–333 (2008).
9. Yan, J. et al. Rare genetic variation at Zea mays crtRB1 increases $\beta$-carotene in maize grain. Nat. Genet. 42, 322–327 (2010).
10. Muthusamy, V. et al. Genetic variability for kernel $\beta$-carotene and utilization of crtRB1 3’TRE gene for biofortification in maize (Zea mays L.). Indian J. Genet. Plant Breed. 72, 189–194 (2012).
11. Babu, R., Rojas, N. P., Gao, S., Yan, J. & Pixley, K. Validation of the effects of molecular marker polymorphisms in LcyE and CrtRB1 on provitamin A concentrations for 26 tropical maize populations. Theor. Appl. Genet. 126, 389–399 (2013).
12. Mertz, E. T., Bates, L. S. & Nelson, O. E. Mutant gene that changes protein composition and increases lysine content of maize endosperm. Science 145, 279–280 (1964).
The authors declare no competing interests.

Author contributions
Conceived and designed the experiments, S.N., H.S.G., F.H.; Performed the experiments, N.C., N.R., B.P., S.C., D.M., A.K., S.V.; Analyzed the data, N.C., S.C., N.R., B.P.; Suggestions, G.K.N., M.S., R.R., V.M.; Writing—review & editing, N.C., A.K., S.C., N.S.

Competing interests
The authors declare no competing interests.
