ISSN 1479-2621

Plant Genetic Resources: Characterization and Utilization (2018) **16**(5); 424–436
doi:10.1017/S1479262118000345

Recombinant inbred lines derived from cultivars of pea for understanding the genetic basis of variation in breeders' traits

Carol Moreau[1], Maggie Knox[1], Lynda Turner[1], Tracey Rayner[1], Jane Thomas[2], Haidee Philpott[2], Steve Belcher[3], Keith Fox[4], Noel Ellis[1,5]† and Claire Domoney[1]†*

[1]*John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK,* [2]*NIAB, Huntingdon Road, Cambridge CB3 0LE, UK,* [3]*Processors and Growers Research Organisation, Great North Road, Thornhaugh PE8 6HJ, UK,* [4]*Limagrain Ltd, Station Road, Docking, King's Lynn PE31 8LS, UK and* [5]*University of Auckland, Auckland, New Zealand*

Received 14 September 2018; Accepted 17 September 2018

Abstract

In order to gain an understanding of the genetic basis of traits of interest to breeders, the pea varieties Brutus, Enigma and Kahuna were selected, based on measures of their phenotypic and genotypic differences, for the construction of recombinant inbred populations. Reciprocal crosses were carried out for each of the three pairs, and over 200 F_2 seeds from each cross advanced to F_{13}. Bulked F_7 seeds were used to generate F_8–F_{11} bulks, which were grown in triplicated plots within randomized field trials and used to collect phenotypic data, including seed weight and yield traits, over a number of growing seasons. Genetic maps were constructed from the F_6 generation to support the analysis of qualitative and quantitative traits and have led to the identification of four major genetic loci involved in seed weight determination and at least one major locus responsible for variation in yield. Three of the seed weight loci, at least one of which has not been described previously, were associated with the marrowfat seed phenotype. For some of the loci identified, candidate genes have been identified. The F_{13} single seed descent lines are available as a germplasm resource for the legume and pulse crop communities.

Keywords: Breeders traits, pea, recombinant inbred lines, seed weight, seed yield

Introduction

As a widely grown pulse crop and one of the oldest domesticated crops, pea (*Pisum sativum* L.) is grown in many regions of the world. The crop has a high content of protein, starch and other nutritional constituents, which make the seeds a valuable source of food and feed and, as a legume, it contributes positively to soil health and so reduces food's environmental impacts (Poore and Nemecek, 2018). Pea breeding has achieved many successes in the development of diverse markets. These include the different uses as a vining crop for fresh and frozen vegetable use, as immature seeds, mangetout and sugar snap pods, as well as the combining crop types that are used for mature seeds, used whole (marrowfat types), or as flour and added ingredients for other foods. As a feed crop, the use of pea is equally diverse, encompassing farm animal and poultry feed and specialist markets for pet and pigeon feed. Additionally, there is renewed interest in developing pea for valuable and healthy wheat-free food products, novel snacks, as well as an alternative to soya for feed formulation.

Despite this interest and need, there are many traits in pea for which their genetic basis is poorly understood

*Corresponding author. E-mail: claire.domoney@jic.ac.uk
†These authors contributed equally to this work

and breeding programmes cannot avail of modern technologies to accelerate crop improvement. Furthermore, there are agronomic traits which require significant improvement for better yield stability in order to promote and sustain a larger growing area. Currently, the key breeding objectives include improving overall yield, yield stability and its components, resistance to biotic and abiotic stresses, as well as enhancing seed quality traits which promote the development of new markets and provide growers with premium returns for their crops. New challenges imposed by climate change, coupled with new regulations regarding seed formulations for disease prevention, are providing additional incentives to crop breeding programmes to diversify the gene pool and to use marker-assisted selection to speed up the introgression of favourable alleles.

Over recent years, many mapping populations have been constructed in pea and deployed to develop genetic maps and identify loci involved in controlling seed and developmental traits (Tayeh *et al.*, 2015*a*, 2015*b*, and citations therein). In many cases, genetic maps were constructed from populations developed from wide crosses, involving diverse germplasm, which delivered an abundance of polymorphic markers and permitted genes of interest to be mapped rapidly and maps to be integrated (Hall *et al.*, 1997*a*, 1997*b*; Laucou *et al.*, 1998, Ellis and Poyser, 2002). In such cases, the populations were not suitable for field study or for the study of agronomic and seed quality traits that are relevant to current agriculture.

In this paper, we investigate the genetic diversity among cultivated pea in comparison with the wider germplasm and choose three contrasting parental lines to generate mapping populations suitable for field trials and in which agronomic traits could be studied. We describe the process by which the parental lines were chosen and report on the identification of major quantitative trait loci for seed size and overall yield.

Materials and methods

Plant materials

A panel of 48 varieties representing pea cultivars which are harvested for dry seed (so-called combining cultivars) was supplied by Limagrain UK Ltd. and the Processors and Growers Research Organisation (PGRO), based on UK National and Recommended Lists (online Supplementary Table S1). Varieties of pea used as a combining crop are generally round- rather than wrinkled-seeded, but with variation for seed shape (block-shaped marrowfat, dimpled), size and colour (green, blue, white/yellow) characteristics, which are related to their end-use (http://www.pgro.org/). A set of 10 diverse pea lines was obtained from the Germplasm Resources Unit at the John Innes Centre (JIC), Norwich, UK. All the cultivated and diverse pea lines used in this study are *Pisum sativum*. Of the diverse lines studied, the most distinct is JI 281, classified as *Pisum sativum* and the accession was collected in Ethiopia (see: https://www.seedstor.ac.uk/search-infoaccession.php?idPlant=23681). Seeds were sown in a glasshouse at JIC and leaves harvested from individual plants for the preparation of DNA.

Reciprocal crosses were carried out between pairs of three chosen variant lines (see below), the cultivars (cv.) Brutus (medium seed size, green cotyledon), Enigma (medium seed size, yellow cotyledon) and Kahuna (large-seeded marrowfat with green cotyledon). The F_1 seeds and plants were verified to be true crosses, and F_2 seeds selfed to generate single seed descent recombinant inbred lines (RILs) to F_{13}. One half of the RILs from each population was derived from one of the two reciprocal crosses between parental lines to give at least 100 RILs per reciprocal cross (>200 RILs per population). The single-seed descent lines generated EK/KE (Enigma × Kahuna and reciprocal), BK/KB (Brutus × Kahuna and reciprocal) and BE/EB (Brutus × Enigma and reciprocal) populations.

Leaves were collected from individual F_6 plants, and leaf DNA used to develop genetic maps for the three populations. Bulked F_7 seeds from the genotyped F_6 plants were multiplied to generate F_8–F_{11} bulks, which were used in field trials alongside the parent lines (6 m^2 plots, 60 plants/m^2) at PGRO and NIAB, Cambridgeshire, UK over the standard growing season (March–July). Seeds were pre-treated with fungicides and trials were protected by cages (NIAB) or other deterrents of predation (PGRO). Single plots of each RIL were grown at F_8 (Year 1, Y1); thereafter, triplicate plots were grown for every RIL (Y2-4 and subset trials below).

Selected RIL bulks were chosen based on contrasting yield over two or more seasons and grown in further trials, using a standard commercial plot size and planting density (18 m^2, 70 plants/m^2). Nineteen RILs were chosen: BE 83, BE 91, EB 114, EB 143, EB 153, EB 173, EK 12, EK 34, EK 48, EK 73, KE 175, KE 180, KE 198, BK 37, BK 63, KB 122, KB 152, KB 193 and KB 201, and grown along with the cv. Prophet as a commercially available high-yielding cultivar.

Trait analysis of the panel of cultivars and RILs

The historical data available for the cultivar panel from National and Recommended List trials of selections from breeding programmes were analysed with respect to priority phenotypic traits: yield, standing ability, downy mildew resistance and seed protein concentration. GGE (genotype and genotype × environment) biplot analysis

https://doi.org/10.1017/S1479262118000345 Published online by Cambridge University Press

(Yan *et al.*, 2000) of the panel, based on a principal component analysis (PCA) of data collected as part of breeders' trials, in combination with the genetic marker analysis (see below), was used to identify three maximally contrasting cultivars as parents for the three-way crosses: the cultivars Brutus (B), Enigma (E) and Kahuna (K).

Traits were scored for RILs and parental lines over all experiments. Consistently, thousand seed weight, overall yield, standing ability, haulm length/plant height and maturity were scored. For standing ability, poor to excellent standing was recorded on a scale of 1–10, according to the procedures for National List trials.

Genetic analysis of the panel of cultivars and the three RIL populations

Analysis of genetic variation among the panel of cultivars in comparison with JI reference pea lines was carried out, using ^{33}P-labelled retrotransposon-based sequence-specific amplified polymorphism (SSAP) genetic markers, which reflect polymorphism of the insertion sites of *Ty1-copia* class retrotransposons, chiefly the *PDR1* retrotransposon (Ellis *et al.*, 1998; Flavell *et al.*, 1998; Jing *et al.*, 2005). A set of diverse pea lines was included in the screen, representing the parents of recombinant inbred mapping populations (JI reference lines: JI 281, JI 15, JI 399, JI 1194, JI 73, JI 1345, JI 1201, JI 813, JI 868 and cv. Birte) as 10 reference lines, which provided highly contrasting genetic backgrounds. Several biological replicates were included in these analyses (see online Supplementary Figs. S1, S2). The marker dataset generated for the cultivar set was analysed using the 'Structure' programme, as described previously (Pritchard *et al.*, 2000; Evanno *et al.*, 2005; Jing *et al.*, 2010).

Genetic markers were developed for the RILs generated from the three chosen lines, using an adaptation of the SSAP marker method above to one based on fluorescently tagged markers, which were analysed using an automated ABI 3730 *xl* platform (Knox *et al.*, 2009). This system provided an improved accuracy of amplicon scoring, increased the available marker number and improved allele discrimination (Knox *et al.*, 2009). The genetic maps developed using SSAP markers were supplemented with gene-specific markers, using available primer sequence information (Page *et al.*, 2002; Aubert *et al.*, 2006). Populations of RILs based on wide crosses (Ellis *et al.*, 2018) were used to investigate the linkage between markers of interest.

Genetic maps were constructed for the three sets of RILs (BE/EB, BK/KB and EK/KE), using JoinMap® 3.0 (Kyazma; Rayner *et al.*, 2017). Quantitative trait scores for RILs were analysed, using interval mapping and MapQTL® (Kyazma) to identify significant genetic marker associations, determined by the logarithm of the odds (LOD) and Kruskal–Wallis significance values.

Results

Selection of parents for generating RILs

The selection of parental lines from the panel of 48 cultivars was based on identifying maximally contrasting lines for both breeders' priority traits and genetic distance, using prior phenotypic data gathered from field trials of the panel of cultivars and genetic marker diversity data, respectively. A GGE biplot analysis (Yan *et al.*, 2000) of the panel, based on a PCA of data collected as part of breeders' trials and relating to phenotype scores for four traits: overall yield, standing ability, downy mildew resistance and seed protein concentration, is shown alongside supporting data in online Supplementary Table S2. Fig. 1 shows an analysis of genotype data for the cultivar set, based on scores for 152 genetic (*PDR1* SSAP) markers. Genotype data were collected (as SSAP marker band presence/absence scores) for the cultivar set plus the JI germplasm reference accessions (designated JI lines), the latter of which included the parents of diverse mapping populations, described previously (Ellis and Poyser, 2002; Vigeolas *et al.*, 2008; Ellis *et al.*, 2018), and included biological replicates for several lines. An example gel used for genotyping is shown in online Supplementary Fig. S1. Fig. 1(a) shows the phylogenetic relationship of the cultivars within the panel in relation to 10 JI reference lines. Most but not all of the JI reference lines are separated and shown at the upper edge of the tree (Fig. 1(a)). The data indicated that the cultivars could be distinguished genetically from each other and clearly from JI 15 and JI 281 (Fig. 1(a)), which represent the very diverse parents of sets of RILs involving JI 15, JI 281, JI 399 and JI 1194 (Hall *et al.*, 1997*a*, 1997*b*; Ellis *et al.*, 2018). The relationship between two JI germplasm genotypes, JI 1194 and JI 1201, should be noted as being closely adjacent. These are two near-isogenic lines (developed by G.A. Marx), with contrasting alleles for three loci that regulate leaf development (*afila, af; stipules-reduced, st, tendril-less, tl*). It is noteworthy that JI 813 lies close to cultivars of the marrowfat class (Fig. 1(a)); JI 813 is derived from the marrowfat cv. Vinco.

The dataset comprising 152 polymorphic markers was used to calculate a distance matrix of (dis)similarity. Compression by principal coordinate analysis (PCO, Fig. 1(b)) showed that at least two major groups of accessions could be distinguished, one of which included marrowfat types (e.g. the cultivars Maro, Princess, Kahuna and Samson, clustered in the right-hand side of the plot). In this plot, the proportion of variance in the first two

https://doi.org/10.1017/S1479262118000345 Published online by Cambridge University Press

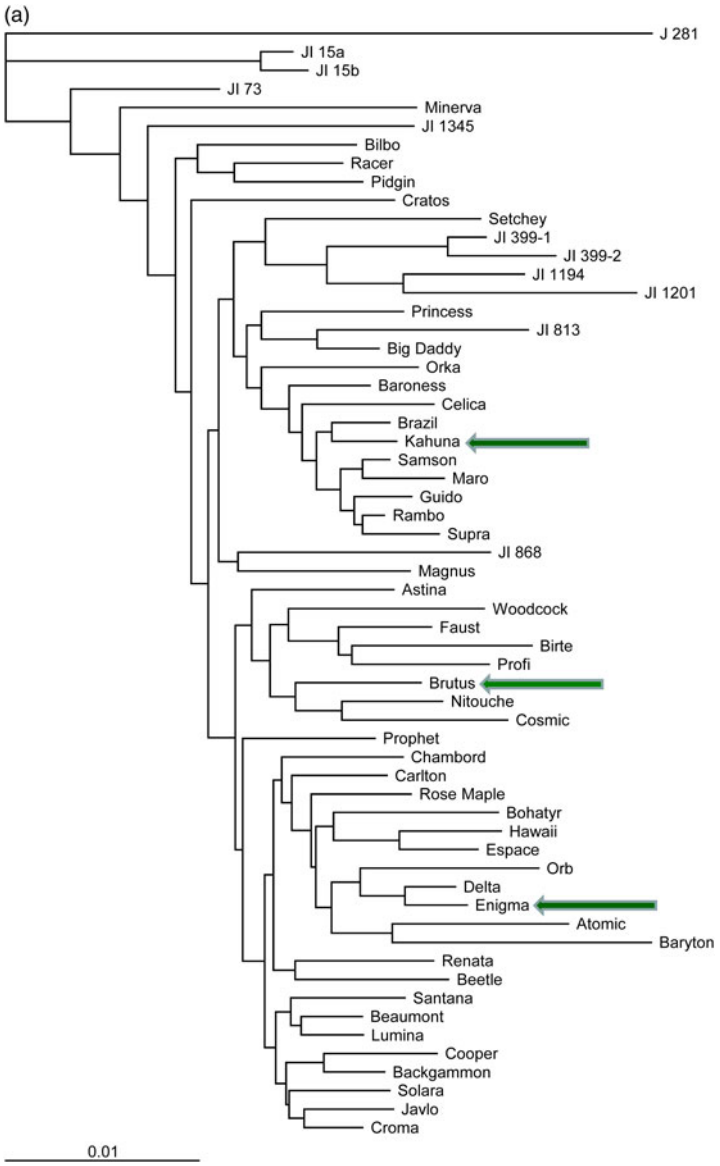

Fig. 1. (a) Phylogenetic relationships among the panel of 48 cultivars in comparison with 10 JI reference lines (JI numbered lines and cv. Birte), based on the genetic analysis of 152 *PDR1* SSAP markers. Where there were marker disagreements between duplicate samples of the same accession, these are numbered separately (JI 15a, b; JI 399-1, -2). The positions adopted by the three cultivars chosen as parents are highlighted by the green arrows. The bar indicates the distance matrix scale, as determined from neighbour-joining phylogenetics. (b). Relationships among the panel of 48 pea cultivars, based on PCO analysis of the genetic marker data, with a projection of the genetic variance data onto planes of the two leading dimensions. The positions adopted by the lines chosen as parents are indicated (black circles). The % variation explained by the two dimensions is indicated (32% for PC1, 29% for PC2).

dimensions is similar and accounts for about 60% of the genetic variation among the cultivars.

Analysis of the marker data obtained for the cultivars, using the population genetics programme 'Structure' (Pritchard *et al.*, 2000; Jing *et al.*, 2010), facilitated a comparison of the chosen parents with the cultivars as a whole (Fig. 2). The 'Structure' programme takes an objective approach to propose common progenitor populations for a given set of genotypes, based on estimations of the number of progenitor populations (K) and their relative contribution to each individual genotype. The value of K is estimated by multiple runs of the programme for different values of K and by investigating how the statistic ln (K|D) varies with K, where ln(K|D) provides an estimate of the likelihood of the data given the modelled K. From the analysis shown in Fig. 2(a), K values of 2, 3 and 4 were investigated further and the correlations of their Q groups are shown (Fig. 2(b)). Fig. 2(c) shows the Q

https://doi.org/10.1017/S1479262118000345 Published online by Cambridge University Press

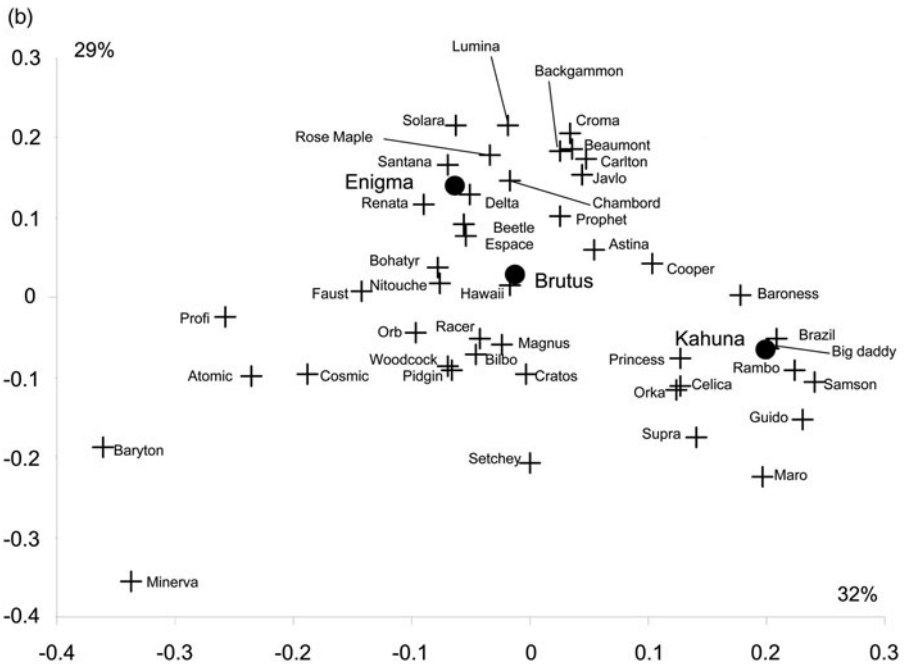

Fig. 1. (Continued).

plots, illustrating the contribution of each presumed progenitor to an individual genotype, for $K=2$, $K=3$ and $K=4$. Although $K=2$ was best supported according to the method of Evanno *et al.* (2005), the $K=3$ plot shows best how the three selected cultivars represent distinct subgroups within the panel. The cultivars with a substantial contribution from sub-population K3,3 (shown in green) are predominantly marrowfat types with some large blues, and all marrowfat lines showed this contribution (Fig. 2(c)).

This 'Structure' analysis (distribution on the Q plots, Fig. 2(c)), together with the marker PCO plot (Fig. 1(b)), was used to select three cultivars that were as distinct as possible on the basis of the genetic marker analysis, while being constrained by also showing contrasting phenotypes (online Supplementary Table S2). In this way, the derived RILs were expected to segregate for traits of interest and to be amenable to genetic analysis. One additional constraint was placed on the final selection of lines: that they should not differ phenotypically because of the allele at the *af* locus since this trait is likely to have major pleiotropic effects that would dominate the characterization of any resulting RIL population. The *afila* (*af*) gene affects leaf morphology (wild-type leafed versus so-called semi-leafless phenotypes) and is likely to be relevant to many agronomic traits, including overall field performance (Burstin *et al.*, 2007); the specific effects of this gene are best investigated in near-isogenic lines. On this basis, the cv. Minerva was ruled out as a parent, even though it is very distinct from most of the recommended list varieties which were analysed (Figs. 1, 2). The lines finally selected as parents (the cultivars Brutus, Enigma and Kahuna) corresponded to different market classes (large blue, white and marrowfat types, respectively) and all were *af* lines. The parental lines are shown to the left of the Q plots in Fig. 2(c) to highlight the relative contribution of their (conjectured) progenitors. The three parents capture 63% of the alleles identified in the cultivars. The frequency of the dominant alleles identified in the three selected lines is strongly correlated with their frequency in cultivars as a whole ($r^2 \sim 0.8$).

In summary, the cvs. Brutus, Enigma and Kahuna were selected as semi-leafless varieties of contrasting market classes for the generation of mapping populations. The parents represented the phenotypic (online Supplementary Table S2) and genotypic (Figs. 1, 2) variation available within the elite pea gene pool. The constraints placed on their selection meant that the parental lines were asymmetrically placed on the phenotypic analysis (online Supplementary Table S2). The consistency of genotype data among the seed lots available for the chosen parental lines was checked, using one SSAP primer combination (online Supplementary Fig. S2).

Establishment of crosses and development of genetic maps

Reciprocal crosses were carried out between pairs of the chosen parents, yielding three populations of RILs, and 220 F_2 seeds (110 for each reciprocal cross) were sown for every cross (Brutus × Enigma, Brutus × Kahuna,

https://doi.org/10.1017/S1479262118000345 Published online by Cambridge University Press

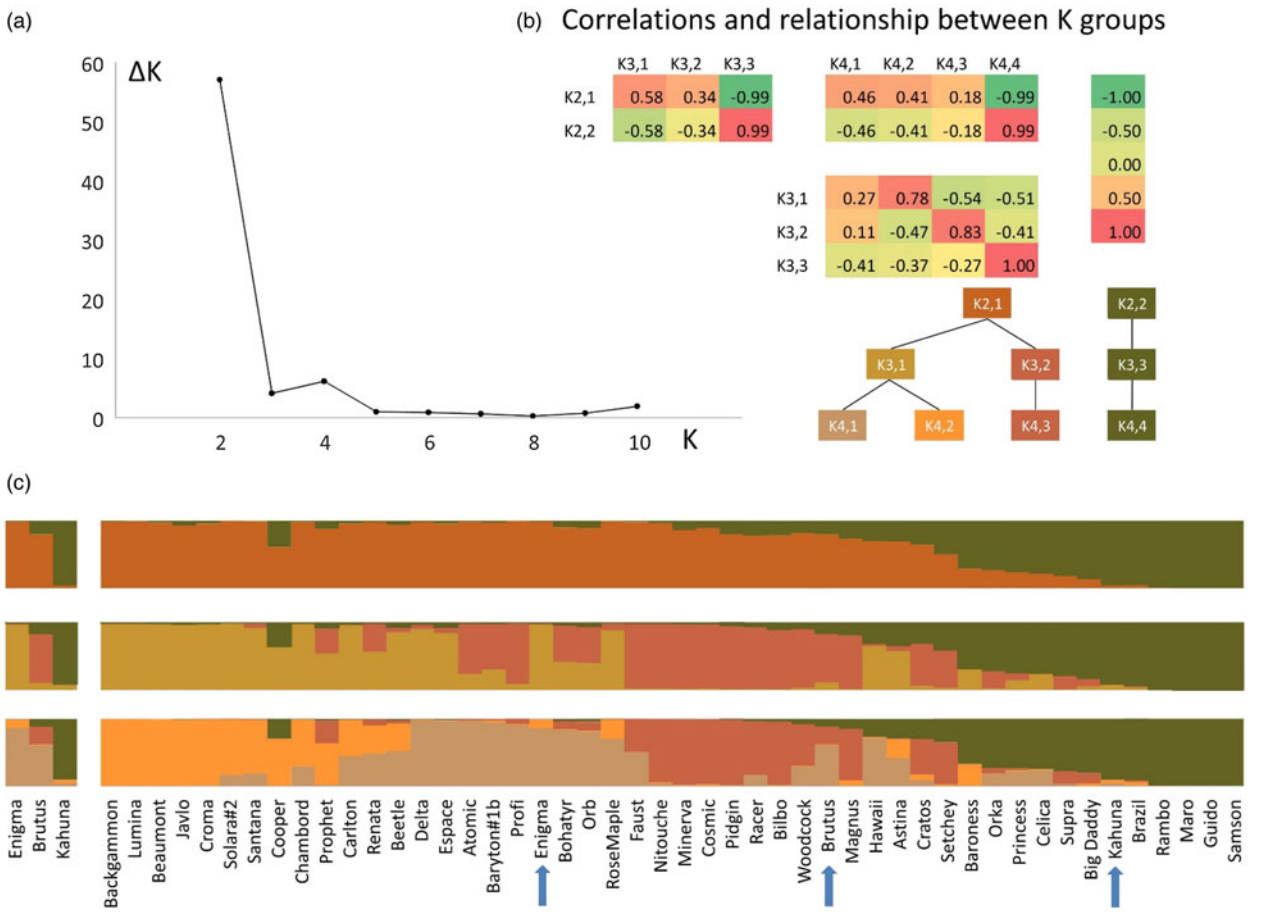

Fig. 2. Structure v 2.1 analysis, based on the genetic marker data obtained for a set of 48 pea cultivars using the default parameter set with the admixture model and comprising 10,000 Markov Chain Monte Carlo (MCMC) runs after a 'burn-in' of 10,000. *K* values in the range 1–10 were examined with five runs for each *K* value. (a) On the basis of the Evanno *et al.* (2005) analysis, the values of $K = 2$, 3 and 4 were investigated. (b) The correlations between *Q* groups within each *K* were calculated in Excel to establish correspondences; non-self correlations between corresponding *Q* groups ranged from $r = 0.993$ to 0.998 for $K = 3$ and $r = 0.987$ to 0.996 for $K = 4$. Within each *K* value, *Q* values for the three most correlated runs were averaged and the correlations between these are presented. From these correlations, the way the groups split as *K* increases has been deduced. (c) The averaged *Q* values for $K = 2$, 3 and 4 are plotted (top to bottom) and each of the varieties is identified. The three parents of the RIL populations are marked with arrows and additionally shown to the left of the *Q* plots.

Enigma × Kahuna as BE/EB, BK/KB and EK/KE RILs, respectively). The three parental lines had contrasting seed traits (yellow or green cotyledon colour; large- or medium-sized seeds). The F_1 seeds and/or plants were checked to prove that they were true hybrids. Hybrid status was confirmed by phenotype (cotyledon colour when Kahuna or Brutus had been the maternal parent, where green cotyledon colour (*i*) is recessive to yellow (*I*) in Enigma, and by genotype, using SSAP marker analysis where markers from both parents were apparent in heterozygous plants; see online Supplementary Figure S2 for parental polymorphisms scored). Online Supplementary Figure S3 (A, B) shows examples of the phenotypes scored for parental and F_1 hybrid seeds, where the cv. Kahuna (a marrowfat) was a parent. The phenotypes of the F_1 seeds obtained for these two crosses indicated that the marrowfat trait may be maternally determined (online Supplementary Fig. S3). The combined results confirmed that the crosses had been successful and allowed the efficient generation of the F_2 populations.

The genetic map data obtained for the three sets of RILs at F_6 are shown in online Supplementary Figs. S4–S6. Alignment of SSAP marker data across diverse populations, including wide crosses, facilitated the map development. Due to the much greater genetic similarity between the cultivar parents, there were as expected far fewer markers available for most linkage groups (LG) in the cultivar-derived RILs than in those derived from wide crosses. It was notable that, in some cases, there was a severe paucity of genetic marker data, potentially indicative of a common origin of chromosomal segments within the

https://doi.org/10.1017/S1479262118000345 Published online by Cambridge University Press

relevant parents. This is particularly true for the BE/EB RILs, where LG IV and VII have two markers each (online Supplementary Fig. S4). In such cases, these common LG regions could be largely discounted as having an association with the control of quantitative traits evident in the derived RILs. In contrast, where the cv. Kahuna is a parent, a much greater number of polymorphic markers was evident for LG VII, in particular (online Supplementary Figs. S5, S6). This possibly indicates much greater distinctness of this LG in the marrowfat class of pea, compared with the other combining varieties.

Trait and quantitative trait locus (QTL) analysis in RILs

At F_8 (F_6 bulks), single plot data were generated for the RILs (year 1, Y1) but, thereafter, triplicate plots were sown for every RIL (Y2-4). Throughout the trials conducted on the three sets of RILs, the principal traits scored were overall yield, standing ability and thousand seed weight. Although susceptibility to downy mildew was additionally considered as a relevant trait to score, this disease was only in evidence to any great extent in year 3 at the PGRO site, where it was associated with generally poor performance due to waterlogging in very wet weather. Equally, standing ability or lodging, a trait that is often scored by its components (creep, followed by erect growth, as opposed to canopy collapse), was not always in evidence. The datasets collected were analysed genetically in two ways: as means of the raw data values and as adjusted data, according to accepted practices for national and recommended list trials, when part plots had been damaged, lost or otherwise affected by non-standard problems, such as invasive weeds.

Fig. 3(a) shows an example of the range of variation for thousand seed weight, as measured in one season (Y4) for EK/KE RILs and parental lines. The low standard error of the mean (SEM, Fig. 3(a)) was typical of measurements for this trait across all populations. Although the range of trait values varied according to the season for all populations (not shown), the parental values fell consistently at either end of the seed weight spectrum, indicating a multigene control and little transgressive segregation (Fig. 3(a)). QTL (quantitative trait locus/loci) analysis of thousand seed weight data revealed a consistent pattern of genetic marker association across years (Table 1, Fig. 3(b)). Two genetic loci were associated with thousand seed weight on LG I: one of these (top of LG I) was apparent when the cv. Kahuna was involved in the cross (BK/KB and EK/KE RILs) and the second (bottom of LG I) was consistent among years for the BE/EB RILs (Fig. 3(b)). The cvs. Kahuna and Enigma contribute positively to the trait at the QTL on the top and bottom of LG I, respectively. Two additional genetic regions were associated with variation in thousand seed weight when Kahuna was a parent, with one of these also apparent in the BE/EB population (Table 1). The QTL on LG IV fell just over the LOD threshold in the BK/KB population (not shown), whereas it was very significant for the EK/KE population (Table 1). A QTL on LG V was evident for two populations, EK/KE and BE/EB (Table 1). Overall, the cv. Kahuna contributed positively to the seed weight trait over three distinct genetic loci, explaining up to 96% of the variation in seed weight (Table 1).

Fig. 4(a) shows an example of the range of variation observed for overall yield in one population. Both parents (cvs. Enigma and Kahuna) showed values at the upper end of the yield spectrum, as would be expected for two commercially cultivated lines, but with an appreciable number of lines showing higher or particularly lower yields than either parent, indicative of transgressive segregation. Although the yield data typically showed much higher SEM values than for other traits (see Fig. 3(a) for example), the maximum yield potential for the RILs was shown to be in excess of 5 t/ha, dependent on the RIL and the season. For overall yield, association with genetic marker data showed more variability, as expected for a complex trait. Nonetheless, some consistency of QTL associations was observed (Table 1, Fig. 4(b)). Two QTL were evident on LG I, with the parent cv. Enigma contributing positively to yield at each locus. One of the loci was on the upper end of LG I (Fig. 4(b)) in a region also associated with thousand seed weight in populations involving cv. Kahuna (Fig. 3(b)). The significance of this yield QTL was enhanced by analysis of data adjusted for non-standard plot effects (Fig. 4(b)). The QTL for yield that was detected towards the lower end of LG I for the BE/EB RILs (Fig. 4(b)) was not coincident with that influencing thousand seed weight in the same cross (Fig. 3(b)). The QTL detected for overall yield on LG III identified a similar region of the LG in all three populations (Fig. 4(b), Table 1). A QTL for yield on LG V was evident in the EK/KE population in 1 year only (Table 1).

A further experiment aimed to establish the components of yield that contributed to the major QTL identified in the three populations. A subset of lines, selected on the basis of relative consistency of yield, was subjected to trials alongside the high-yielding cultivar, Prophet, using commercially-relevant plot size and sowing density. A very strong correlation ($R^2 = 0.92$) between overall yield and standing ability was apparent in one such trial, where some lines (including cv. Prophet) yielded in excess of 5 t/ha (online Supplementary Fig. S7).

Candidate genes for traits

The genetic location of some of the QTL for thousand seed weight data in this work prompted an investigation into candidate genes within the genetic regions identified.

https://doi.org/10.1017/S1479262118000345 Published online by Cambridge University Press

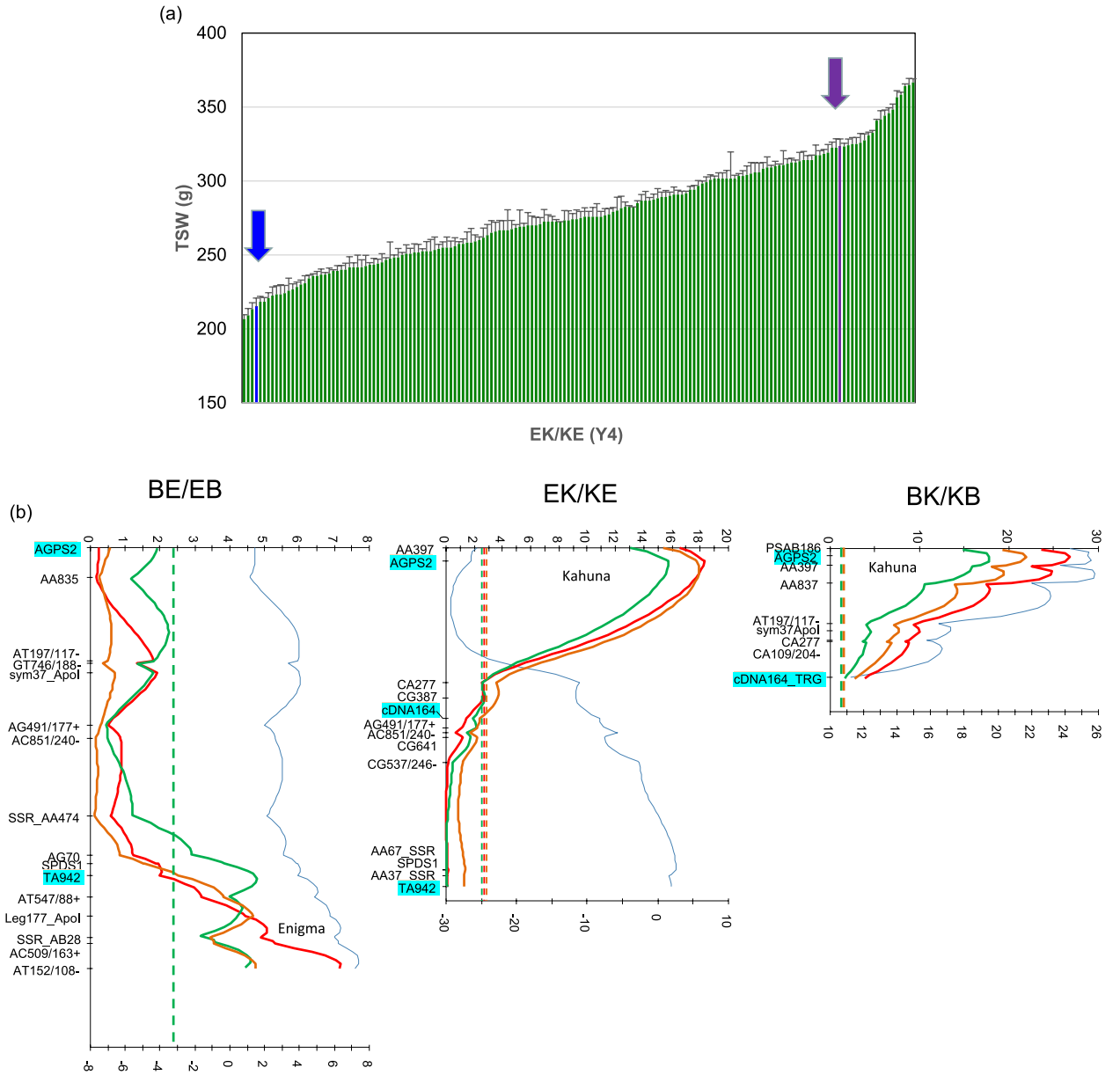

Fig. 3. (a) An example of the range of variation for thousand seed weight (TSW), as measured for the EK/KE population (year 4, triplicate plots of 166 lines, PGRO site). The positions of the parental lines are indicated by arrows (blue, cv. Enigma; purple, cv. Kahuna) with values of 215.0 ± 5.8 and 323.3 ± 5.1 (mean ± SE), respectively. (b) Major QTL on linkage group (LG) I for TSW in three RIL populations across 3 years (red, Y1; green, Y2; brown, Y4). Peaks are shown above the LOD threshold (dotted vertical lines, top scale; LOD 2.6-2.7 for BK/KB, 2.4 for BE/EB, 2.5–2.8 for EK/KE) at the bottom of LG I for BE/EB and at the top of LG I for EK/KE and BK/KB populations. The additive genetic effect is shown (blue line, bottom scale), with the parent contributing positively to the trait indicated in every case. The linkage groups are aligned, using genetic markers in common within the populations analysed (blue highlight) and with additional wide crosses.

This included two candidates, *AgpS2* (Fig. 3b) and *subtilisin*, the latter of which mapped close to *af* (leaf phenotype) on the lower end of LG I in additional crosses (cv. Princess × JI 185, not shown) and to the syntenic region of chromosome 5 in *Medicago truncatula* (D'Erfurth *et al.*, 2012).

The predicted amino acid sequences for the small subunit 2 of ADP-glucose pyrophosphorylase gene (*AgpS2*) in 10 pea lines, including the three parental lines from this study, revealed one amino acid difference in cv. Kahuna, compared with other lines (online Supplementary Fig. S8). Although this substitution might

https://doi.org/10.1017/S1479262118000345 Published online by Cambridge University Press

Table 1. Summary of QTL identified in three populations (BE/EB, EK/KE, BK/KB) for thousand seed weight (TSW) and yield traits over 4 years

Population	Trait	Year	Maximum LOD	LOD threshold	% variation	Parent (positive)	LG	Genetic marker
BE/EB	TSW	1	7	2.4	19	Enigma	I	AT152/108-
BE/EB	TSW	2	5	2.4	13	Enigma	I	AT152/108-
BE/EB	TSW	4	5	2.4	12	Enigma	I	Leg177_ApoI
EK/KE	TSW	1	18	2.7	42	Kahuna	I	AgpS2
EK/KE	TSW	2	16	2.5	37	Kahuna	I	AgpS2
EK/KE	TSW	4	18	2.8	42	Kahuna	I	AgpS2
BK/KB	TSW	1	27	2.6	51	Kahuna	I	AgpS2
BK/KB	TSW	2	12	2.6	26	Kahuna	I	AgpS2
BK/KB	TSW	4	22	2.7	43	Kahuna	I	AgpS2
EK/KE	TSW	1	4	2.7	13	Kahuna	IV	AC214/168-
EK/KE	TSW	2	4	2.5	18	Kahuna	IV	AT975
EK/KE	TSW	4	5	2.8	20	Kahuna	IV	AT975
BE/EB	TSW	1	3	2.4	10	Brutus	V	TA519
BE/EB	TSW	4	4	2.4	14	Brutus	V	TA519
EK/KE	TSW	1	4	2.7	20	Kahuna	V	GC327/386-
EK/KE	TSW	2	4	2.5	17	Kahuna	V	TA519
EK/KE	TSW	4	5	2.8	25	Kahuna	V	GC327/386-
BE/EB	Yield	2	5	2.4	13	Enigma	I	TA942
BE/EB	Yield	3	5	2.4	14	Enigma	I	AG70
EK/KE	Yield	4	3	2.5	12	Enigma	I	AgpS2
BE/EB	Yield	3	3	2.4	9	Enigma	III	AA668/141-
BE/EB	Yield	4	9	2.5	27	Enigma	III	La_Della
EK/KE	Yield	2	3	2.6	7	Enigma	III	TT496
BK/KB	Yield	2	3	2.4	11	Brutus	III	AT513
BK/KB	Yield	3	3	2.5	12	Brutus	III	AT513
EK/KE	Yield	4	3	2.5	15	Enigma	V	TG288/40-

The maximum peak LOD scores, LOD threshold, % variation explained by the locus, the parental line contributing positively to the trait, linkage group (LG) and close genetic markers are listed for the trait QTL.

be significant (K454I), it is not present in a second marrowfat line, cv. Princess, included in the analysis.

Although genetic variation in *subtilisin* has been associated with significant differences in mean seed weight in two legume species (D'Erfurth *et al.*, 2012), the gene sequences determined for the entire coding region of *subtilisin* in cvs. Brutus, Enigma and Kahuna (2290 bp) showed no nucleotide polymorphisms. Some few polymorphisms were apparent in comparisons with two additional lines (JI 281, JI 185; not shown).

The association between yield and genetic markers on LG III identified the region containing the 'La Della' gene marker as being of interest. The marker is based on the gene encoding the pea putative gibberellin (GA) signalling DELLA protein LA (GenBank: DQ848351.1; Weston *et al.*, 2008). This genetic region in the middle of LG III also contains the marker A001 associated with lodging resistance (a component of standing ability) (Tar'an *et al.*, 2003; 2004). The linkage between the A001 and La Della markers was checked in a wide mapping population (JI 15 × JI 399) and three recombinants were identified out of 85 lines scored.

Discussion

In this work, we generated and used three populations of RILs from crosses of cultivated lines of pea to gain an

https://doi.org/10.1017/S1479262118000345 Published online by Cambridge University Press

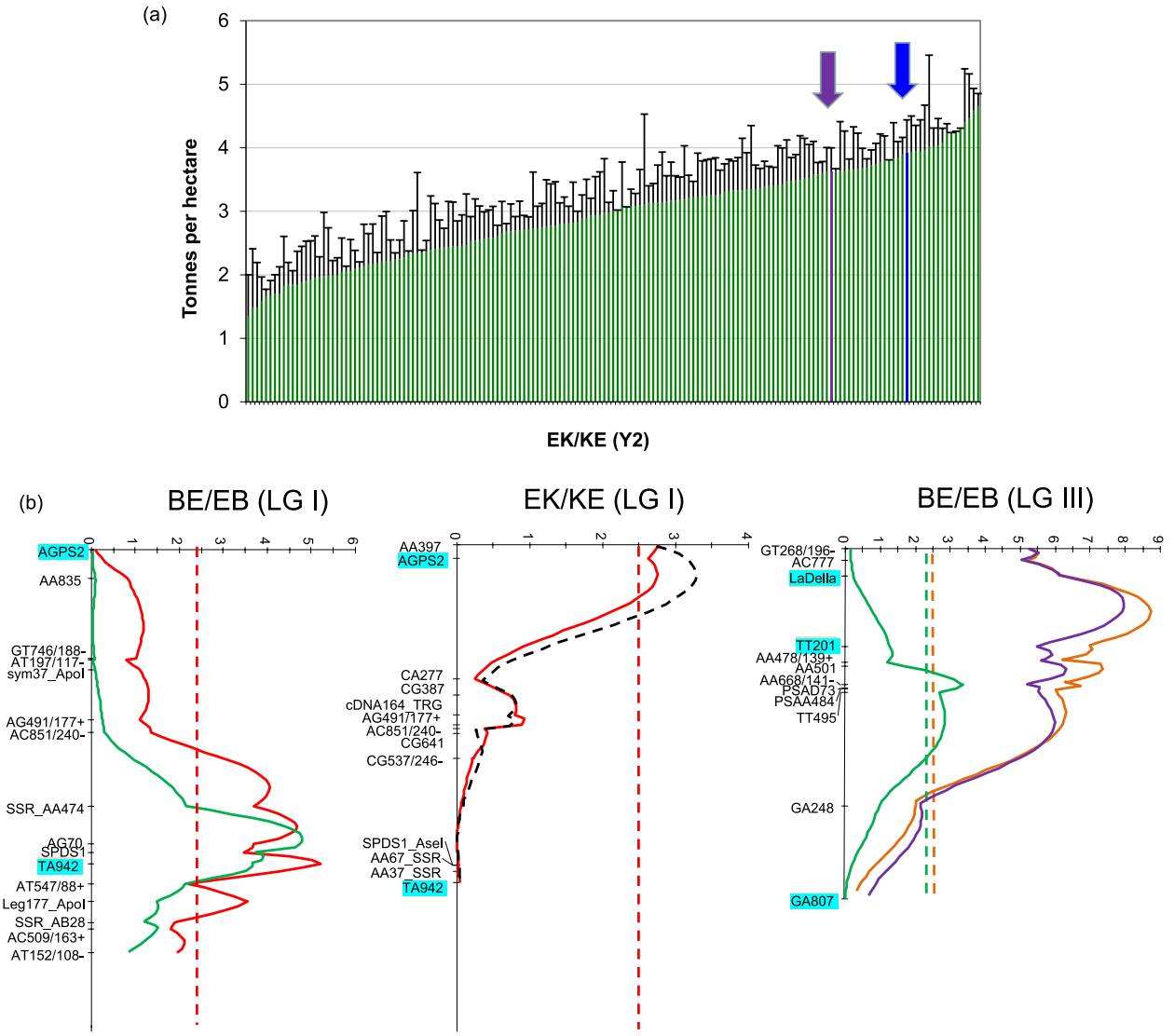

Fig. 4. (a) An example of the range of variation in yield, as determined for the EK/KE population (year 2; triplicate plots of 166 lines, for which triplicate plot data were returned for 159; PGRO site). The positions of the parental lines are indicated by arrows (blue, cv. Enigma; purple, cv. Kahuna) with values of 3.923 ± 0.52 and 3.627 ± 0.37 t/ha (mean ± SE), respectively. (b) QTL on linkage groups I and III (LG I, LG III) for yield in two RIL populations across one (EK/KE; red, Y4 raw; black dashes, Y4 adjusted data) or two (BE/EB LG I; red, Y2; green, Y3; BE/EB LG III; green, Y3; brown, Y4 raw; purple, Y4 adjusted) years. Adjusted data are corrected for areas of plots affected by non-standard influences. Peaks are shown above the LOD threshold (dotted vertical lines, top scale; LOD 2.4 for BE/EB (LG I), 2.5 for EK/KE (LG I), 2.4–2.5 for BE/EB LG III). The parent cv. Enigma contributed positively to the trait for each QTL. The linkage groups can be aligned, using genetic markers in common within the populations analysed (blue highlight) and with additional wide crosses.

understanding of the genetic basis for traits which are relevant to the agronomic and economic performance of the pea crop. Within the limits of the genetic background of cultivated crops, the three parents were chosen to have contrasting genotypes and phenotypes, the former according to genetic marker analysis and the latter according to available commercial trial data for agronomically important traits. The parents and hence the RILs also showed contrasting seed size, a trait of economic importance, with the large block-shaped and somewhat dimpled form of a marrowfat pea seeds being desirable for a variety of food uses. The RILs provide a resource that is available for the mapping of further traits not analysed here, such as seed composition and disease resistance.

The QTL identified for thousand seed weight included two loci on LG I (Fig. 3(b)), one of which has not been described previously and was associated with the large-seeded marrowfat trait of cv. Kahuna. The *AgpS2* gene

https://doi.org/10.1017/S1479262118000345 Published online by Cambridge University Press

in this region might be considered a strong candidate gene for seed size, due to the role of AgpS2 as a subunit of plastidial ADP-glucose pyrophosphorylase, a key regulatory enzyme of starch biosynthesis, which provides a substrate for starch synthase (Weigelt *et al.*, 2009). Furthermore, the small subunits of this enzyme have been shown to play a regulatory role in determining its overall activity through dimerization (Hädrich *et al.*, 2012). However, no consistent amino acid differences were predicted for the two marrowfat lines in comparison with the others analysed in this work. It is possible that differences in the promoter or additional non-coding sequences influence the expression of this gene, which would be expected to impact on seed development. On the other hand, orthologues of transcription regulators such as BS1 in *Medicago truncatula* and *Glycine max,* which when down-regulated led to significant increases in seed size (Ge *et al.*, 2016), may reside at this (or other) QTL identified for thousand seed weight in pea; based on considerations of synteny alone, BS1 (*Medicago* chromosome 1, syntenic to pea LG II) is not a likely candidate.

The QTL for thousand seed weight on the bottom of LG I (Fig. 3(b)) may be explained by variation in the expression levels or pattern of subtilase/subtilisin, previously reported to affect seed size in induced mutants of *Medicago truncatula* and pea (D'Erfurth *et al.*, 2012). No polymorphisms were detected for this protein among the parents used in the present study. In the study of D'Erfurth *et al.* (2012), an association between variation in this gene and ecotypes of both species was reported, although the nucleotide polymorphism associated with the trait in pea did not lead to an amino change in the protein (G612A; K204 K). The substrates for specific subtilase/subtilisin-like proteases are largely unknown, although some are likely to be involved in the maturation of peptide hormones (Srivastava *et al.*, 2008). For the remaining QTL for thousand seed weight (Table 1), the paucity of markers prevented the identification of associated candidate genes of interest. Other authors have reported QTL for seed weight in pea, involving all LG except LG II (Timmerman-Vaughan *et al.*, 1996; Burstin *et al.*, 2007). The LG IV locus identified here (Table 1) may provide a link between these different studies. The LG I locus identified by Burstin *et al.* (2007) may be equivalent to that identified in the BE/EB population at the lower end of the LG (Fig. 3(b), Table 1). Although a marrowfat line was used as a parent in the study of Timmerman-Vaughan *et al.* (1996), a QTL for seed weight was not detected on LG I, possibly reflecting a low density of genetic markers.

The genetic regions associated with yield (Fig. 4(b)) included two QTL on LG I. Although one of these was detected in one population in one year only, it is notable in that it likely corresponds to the same region linked with thousand seed weight when the cv. Kahuna is a parent (Fig. 3(b)). This may indicate a trade-off between seed size and yield under some environmental conditions; here the cv. Enigma promoted higher yield, whereas the cv. Kahuna promoted a higher thousand seed weight (Figs. 3, 4). The QTL for yield on the lower region of LG I (Fig. 4(b)) is not coincident with that for thousand seed weight but its proximity to this QTL in the same population (BE/EB; Fig. 3(b)) and furthermore to genetic loci which control cotyledon colour (*sgr*) and leaf shape (*af*) (Burstin *et al.*, 2007; 2015) might suggest that selection within breeding programmes for seed traits, such as size and colour, and leaf traits could result in counter-selection against overall yield.

The region of LG III associated with overall yield is of particular interest, due to the proximity of two genetic markers designated 'La Della' (Fig. 4(b)) and 'A001', the latter of which has been associated with lodging resistance in the work of Tar'an *et al.*, (2003, 2004). This QTL does not appear to correspond to one reported previously for yield at the lower end of LG III (Burstin *et al.*, 2007). Although the identity of the gene corresponding to the marker A001 remains unknown, it maps in the region of LG III where the internode length-determining gene *la* is located (Ellis and Poyser, 2002; Tar'an *et al.*, 2003; 2004), but correspondence between *la* and either of these genetic markers has not been demonstrated. The recessive alleles *la* and *cry*^{s} act together to confer a long-internode 'slender' phenotype (Potts *et al.*, 1985) and thus may be candidates for *GAI* homologues, where GAI expression inhibits the growth of plants, an inhibition which is antagonized by GA. The 'La Della' marker corresponds to the putative GA signalling DELLA protein LA (Weston *et al.*, 2008). These authors suggest that the *LA* and *CRY* genes encode DELLA proteins, previously characterized in other species (*Arabidopsis thaliana* and several grasses) as repressors of growth and that the action of these genes is destabilized by GA. The role of DELLA proteins in GA signalling pathways, as negative regulators of GA function and their association with 'green revolution' genes (Serrano-Mislata *et al.*, 2017), provides a useful lead in unravelling this genetic locus. Altered expression of *GAI* or *gai* genes in plants can result in tall or dwarfed plants. Generally, dwarf plants are useful in reducing crop losses due to lodging. The demonstration of the strong relationship between yield and standing ability for the subset of RILs tested in this work under commercially-relevant field conditions provides further support for a detailed analysis of this locus in pea.

In this study, we provide useful genetic markers for thousand seed weight and overall yield traits in pea. Although amino acid variation consistent with differences in the seed weight trait was not revealed for the candidate genes identified, further analysis is needed to examine relative expression levels of these genes during seed development. It is possible that some of the candidate genes

https://doi.org/10.1017/S1479262118000345 Published online by Cambridge University Press

identified here will provide perfect markers for the traits being studied, in particular, for yield and standing ability. Although the SSAP markers used throughout this work are not readily transferable, they have provided a cost-effective method to identify genetic loci of interest in specific populations and have demonstrated the utility of the resource described here. The data presented will be developed within a detailed analysis of the loci identified, based on using the forthcoming single nucleotide polymorphism platforms to develop high-density genetic maps (Duarte *et al.*, 2014; Tayeh *et al.*, 2015*b*) in the more advanced RILs (F_{13}). The three inter-related populations of RILs generated provide ideal material for this further research and will be made available through the John Innes Centre Germplasm Resources Unit, UK.

Supplementary material

The supplementary material for this article can be found at https://doi.org/10.1017/S1479262118000345

Acknowledgements

This work was supported by Biotechnology and Biological Sciences Research Council (BBSRC) (BB/J004561/1 and BB/P012523/1) and the John Innes Foundation, and the Department for Environment, Food and Rural Affairs (Defra) (CH0103 and CH0110, Pulse Crop Genetic Improvement Network). We are extremely grateful to Dr Jitender Cheema, JIC, for assistance with quantitative genetic analysis. We are grateful to Hilary Ford and Lionel Perkins, JIC, for their horticultural expertise and management of the recombinant inbred populations. We thank Barrie Smith and Rob Glover, PGRO, for assistance with the early stage field trials. We thank Mike Ambrose, JIC, for developing advanced single seed descent lines of the mapping populations as a bulked germplasm resource.

References

Aubert G, Morin J, Jacquin F, Loridon K, Quillet MC, Petit A, Rameau C, Lejeune-Hénaut I, Huguet T and Burstin J (2006) Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume *Medicago truncatula*. *Theoretical and Applied Genetics* 112: 1024–1041.

Burstin J, Marget P, Huart M, Moessner A, Mangin B, Duchene C, Desprez B, Munier-Jolain N and Duc G (2007) Developmental genes have pleiotropic effects on plant morphology and source capacity, eventually impacting on seed protein content and productivity in pea. *Plant Physiology* 144: 768–781.

Burstin J, Salloignon P, Chabert-Martinello M, Magnin-Robert J-B, Siol M, Jacquin F, Chauveau A, Pont C, Aubert G, Delaitre C, Truntzer C and Duc G (2015) Genetic diversity and trait genomic prediction in a pea diversity panel. *BMC Genomics* 16: 105.

D'Erfurth I, Le Signor C, Aubert G, Sanchez M, Vernoud V, Darchy B, Lherminier J, Bourion V, Bouteiller N, Bendahmane A, Buitink J, Prosperi JM, Thompson R, Burstin J and Gallardo K (2012) A role for an endosperm-localized subtilase in the control of seed size in legumes. *New Phytologist* 196: 738–751.

Duarte J, Rivière N, Baranger A, Aubert G, Burstin J, Cornet L, Lavaud C, Lejeune-Hénaut I, Martinant J-P, Pichon J-P, Pilet-Nayel M-L and Boutet G (2014) Transcriptome sequencing for high throughput SNP development and genetic mapping in pea. *BMC Genomics* 15: 126.

Ellis THN and Poyser SJ (2002) An integrated and comparative view of pea genetic and cytogenetic maps. *New Phytologist* 153: 17–25.

Ellis THN, Poyser SJ, Knox MR, Vershinin AV and Ambrose MJ (1998) Ty1-copia class retrotransposon insertion site polymorphism for linkage and diversity analysis in pea. *Molecular and General Genetics* 260: 9–19.

Ellis N, Hattori C, Cheema J, Donarski J, Charlton A, Dickinson M, Venditti G, Kaló P, Szabó Z, Kiss G and Domoney C (2018) NMR metabolomics defining genetic variation in pea seed metabolites. *Frontiers in Plant Science* 9: 1022.

Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620.

Flavell AJ, Knox MR, Pearce SR and Ellis THN (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. *The Plant Journal* 16: 643–650.

Ge L, Yu J, Wang H, Luth D, Bai G, Wang K and Chen R (2016) Increasing seed size and quality by manipulating BIG SEEDS1 in legume species. *Proceedings of the National Academy of Sciences USA* 113: 12414–12419.

Hädrich N, Hendriks JHM, Kötting O, Arrivault S, Feil R, Zeeman SC, Gibon Y, Schulze WX, Stitt M and Lunn JE (2012) Mutagenesis of cysteine 81 prevents dimerization of the APS1 subunit of ADP-glucose pyrophosphorylase and alters diurnal starch turnover in *Arabidopsis thaliana* leaves. *The Plant Journal* 70: 231–242.

Hall KJ, Parker JS and Ellis THN (1997*a*) The relationship between genetic and cytogenetic maps of pea. I. Standard and translocation karyotypes. *Genome* 40: 744–754.

Hall KJ, Parker JS, Ellis THN, Turner L, Knox MR, Hofer JMI, Lu J, Ferrandiz C, Hunter PJ, Taylor JD and Baird K (1997*b*) The relationship between genetic and cytogenetic maps of pea. II. Physical maps of linkage mapping populations. *Genome* 40: 755–769.

Jing R, Knox MR, Lee JM, Vershinin AV, Ambrose M, Ellis THN and Flavell AJ (2005) Insertional polymorphism and antiquity of PDR1 retrotransposon insertions in *Pisum* species. *Genetics* 171: 741–752.

Jing R, Vershinin A, Grzebyta J, Shaw P, Smýkal P, Marshall D, Ambrose MJ, Ellis THN and Flavell A (2010) The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. *BMC Evolutionary Biology* 10: 44.

Knox MR, Moreau C, Lipscombe J and Ellis THN (2009) High-throughput retrotransposon-based fluorescent markers: improved information content and allele discrimination. *Plant Methods* 5: 10.

Laucou V, Haurogné K, Ellis N and Rameau C (1998) Genetic mapping in pea. *1- RAPD-based genetic linkage map*

of Pisum sativum. *Theoretical and Applied Genetics* 97: 905–915.

Page D, Aubert G, Duc G, Welham T and Domoney C (2002) Combinatorial variation in coding and promoter sequences of genes at the *Tri* locus in *Pisum sativum* accounts for variation in trypsin inhibitor activity in seeds. *Molecular Genetics and Genomics* 267: 359–369.

Poore J and Nemecek T (2018) Reducing food's environmental impacts through producers and consumers. *Science* 360: 987–992.

Potts WC, Reid JB and Murfet IM (1985) Internode length in Pisum. Gibberellins and the slender phenotype. *Physiologia Plantarum* 63: 357–364.

Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Rayner T, Moreau C, Ambrose M, Isaac PG, Ellis N and Domoney C (2017) Genetic variation controlling wrinkled seed phenotypes in *Pisum*: how lucky was Mendel? *International Journal of Molecular Sciences* 18: 1205.

Serrano-Mislata A, Bencivenga S, Bush M, Schiessl K, Boden S and Sablowski R (2017) DELLA genes restrict inflorescence meristem function independently of plant height. *Nature Plants* 3: 749–754.

Srivastava R, Liu JX and Howell SH (2008) Proteolytic processing of a precursor protein for a growth-promoting peptide by a subtilisin serine protease in Arabidopsis. *The Plant Journal* 56: 219–227.

Tar'an B, Warkentin T, Somers DJ, Miranda D, Vandenberg A, Blade S, Woods S, Bing D, Xue A, DeKoeyer D and Penner G (2003) Quantitative trait loci for lodging resistance, plant height and partial resistance to mycosphaerella blight in field pea (*Pisum sativum* L.). *Theoretical and Applied Genetics* 107: 1482–1491.

Tar'an B, Warkentin T, Somers DJ, Miranda D, Vandenberg A, Blade S and Bing D (2004) Identification of quantitative trait loci for grain yield, seed protein concentration and maturity in field pea (*Pisum sativum* L.). *Euphytica* 136: 297–306.

Tayeh N, Aluome C, Falque M, Jacquin F, Klein A, Chauveau A, Bérard A, Houtin H, Rond C, Kreplak J, Boucherot K, Martin C, Baranger A, Pilet-Nayel M-L, Warkentin TD, Brunel D, Marget P, Le Paslier M-C, Aubert G and Burstin J. (2015*a*) Development of two major resources for pea genomics: the GenoPea 13.2 K SNP array and a high-density, high-resolution consensus genetic map. *The Plant Journal* 84: 1257–1273.

Tayeh N, Aubert G, Pilet-Nayel M-L, Lejeune-Hénaut I, Warkentin TD and Burstin J (2015*b*) Genomic tools in pea breeding programs: status and perspectives. *Frontiers in Plant Science* 6: 1037.

Timmerman-Vaughan GM, McCallum JA, Frew TJ, Weeden NF and Russel AC (1996) Linkage mapping of quantitative trait loci controlling seed weight in pea (*Pisum sativum* L.). *Theoretical and Applied Genetics* 93: 431–439.

Vigeolas H, Chinoy C, Zuther E, Blessington B, Geigenberger P and Domoney C (2008) Combined metabolomic and genetic approaches reveal a link between the polyamine pathway and albumin 2 in developing pea seeds. *Plant Physiology* 146: 74–82.

Weigelt K, Küster H, Rutten T, Fait A, Fernie AR, Miersch O, Wasternack C, Emery RJN, Desel C, Hosein F, Martin Müller Saalbach I and Weber H (2009) ADP-glucose pyrophosphorylase-deficient pea embryos reveal specific transcriptional and metabolic changes of carbon-nitrogen metabolism and stress responses. *Plant Physiology* 149: 395–411.

Weston DE, Elliott RC, Lester DR, Rameau C, Reid JB, Murfet IC and Ross JJ (2008) The pea DELLA proteins LA and CRY are important regulators of gibberellin synthesis and root growth. *Plant Physiology* 147: 199–205.

Yan W, Hunt LA, Sheng Q and Szlavnics Z (2000) Cultivar evaluation and mega-environment investigation based on GGE biplot. *Crop Science* 40: 597–605.