

# Seed yield in perennial ryegrass (*Lolium perenne* L.): comparative importance of component traits and detection of seed-yield-related QTL

A. M. Sartie · H. S. Easton · C. Matthew · M. P. Rolston · M. J. Faville

Received: 30 April 2018 / Accepted: 12 November 2018 / Published online: 16 November 2018  
© Springer Nature B.V. 2018

**Abstract** The aims of our study were to evaluate relationships amongst morphological traits associated with seed production in a perennial ryegrass biparental population and to identify genomic regions associated with phenotypic variation in those traits using QTL analysis. This was achieved using data from two field experiments at Palmerston North and Lincoln, New Zealand, in 2003, and days to heading (DTH), reassessed in 2004. Trait association was determined for the Palmerston North experiment where measured traits included seed yield per plant (SYPlant), seed yield per spike (SYSp), reproductive tiller number

(RTiller), spikelets per spike (SpktSp), florets per spikelet (FSpkt), 1000 seed weight (TSW), spike length (SpLen), florets per spike (FSp), floret site utilization (FSUtil), spread of heading (SOH) and plant growth habit (PGHabit). Traits contributing to SYPlant in order of descending value were FSpkt, FSUtil, and RTiller. High TSW was only weakly linked to SYPlant. FSUtil, SOH and RTiller were identified as valuable breeding targets for improving seed yield potential in perennial ryegrass. QTL were identified for all traits except for RTiller. QTL for SYPlant occurred on linkage groups (LG) 2 and 6. Both were co-located with QTL for SYSp and sets of SYPlant components or related traits (FSpkt, FSp; FSUtil and TSW). Major QTL for DTH were identified on LG2 and LG4 and minor QTL on LG7 in consecutive years. There was a strong genotype-by-environment interaction for SYPlant that was reflected in a lack of consistent QTL across environments, while QTL for SYSp and DTH were stable across environments. Identification of component traits and QTL important for seed yield may accelerate genetic improvement in perennial ryegrass through conventional and marker-assisted breeding, respectively.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10681-018-2307-z>) contains supplementary material, which is available to authorized users.

A. M. Sartie · H. S. Easton · M. J. Faville (✉)  
Grasslands Research Centre, AgResearch Ltd.,  
Private Bag 11008, Palmerston North 4442, New Zealand  
e-mail: marty.faville@agresearch.co.nz

A. M. Sartie · C. Matthew  
School of Agriculture and Environment, Massey  
University, Private Bag 11222, Palmerston North 4442,  
New Zealand

A. M. Sartie  
International Livestock Research Institute, Addis Ababa,  
Ethiopia

M. P. Rolston  
Lincoln Research Centre, AgResearch Ltd.,  
Private Bag 4749, Christchurch 8140, New Zealand

**Keywords** Component trait association · Heading date · Marker-assisted breeding · Perennial ryegrass · QTL · Seed yield improvement

## Introduction

Perennial ryegrass (*Lolium perenne* L.) is one of the most widely sown forage grass species in the world, and is the main source of energy and protein for grazing livestock in higher fertility pastures in New Zealand. The development of cultivars with potential for increased productivity is a key objective in forage improvement.

Seed yield is a measure of the total saleable seeds recovered from harvesting and processing a seed crop. An economic level of seed yield is vital to the production and successful commercialisation of any forage grass. However, seed yield in perennial ryegrass, as in other forage grasses, has received less attention from plant breeders than herbage production and forage quality (Marshall and Wilkins 2003), because these species are primarily used as forage.

Seed yield is a quantitative trait, with its expression influenced by many genes and by the environment (Casler et al. 2003). Identification of and selection for relevant component traits is a means to enhance genetic gain for complex, quantitative traits in plant breeding (Donald 1968; Sparnaaij and Bos 1993). Such an approach dissects complex traits into components under separate and probably simpler genetic control, providing useful information on the genetic architecture and insights on the physiological control of the overarching trait. Improved genetic gain for the complex trait may subsequently be achievable by selecting on positively correlated components with sufficient heritability. Based on glasshouse and field studies, component traits with significant contributions to seed yield in perennial ryegrass have been identified. These include increased seed-set (higher proportion of ovules forming seeds), greater seed number per tiller and more reproductive tillers per plant (Marshall and Wilkins 2003; Studer et al. 2008). However, in these studies, only a few traits were measured. Therefore, a more comprehensive analysis of trait associations among traits contributing to seed yield would be very timely.

As a molecular strategy for accelerating trait improvement in plants (Heffner et al. 2009; Hayes et al. 2013), genomic selection (Meuwissen et al. 2001), underpinned by low cost single nucleotide polymorphism (SNP) marker systems (Elshire et al. 2011), is poised to supersede marker-assisted selection that uses molecular markers linked to quantitative trait

loci (QTL) (e.g. Barrett et al. 2008). However, QTL analysis remains an effective tool for deciphering the genetic architecture of quantitative traits, identifying genetic interactions and inter-dependencies amongst traits, and providing a platform for fine mapping and identification of candidate genes.

Regardless of the state of play of genetic and bioinformatics technology, the desirable phenotype to be bred for still needs to be identified. Seed yield is a complex trait, and the identification of its component traits and their interactions, and related traits and the genetic factors controlling them, remains only partly investigated despite a number of insightful studies in the last 10 years, including, among others, Armstead et al. (2008), Byrne et al. (2009), and Brown et al. (2010). For example, the first mentioned study primarily investigated QTL for seed set, the second study investigated a group of three coincident QTL for days to heading, spike length and spikelets per spike, while the third was confined to spike description traits, without investigating threshed seed yield.

Considering the above, the aims of this research were (i) to evaluate relationships amongst traits associated with seed yield potential in a New Zealand-adapted perennial ryegrass population; and (ii) to identify genomic regions associated with phenotypic variation in those traits via QTL discovery and comparative analysis. In this paper we present a comprehensive analysis of associations amongst 13 seed-yield-related traits in the  $I \times S$  mapping population, (Sartie et al. 2011) measured at one location (Palmerston North, New Zealand), and identification of putative QTL for those traits. Concurrent assessment of four traits in the same mapping population at a second location (Lincoln, New Zealand), enabled identification of environmentally stable QTL for SYSp and DTH and, for the latter, further validation was afforded by repeated QTL detection in consecutive years.

## Materials and methods

### Plant material

Two hundred progeny of a previously-described perennial ryegrass mapping population,  $I \times S$  (Sartie et al. 2011), and the two parents were assessed for seed yield potential. Plants of mapping population  $I \times S$

are F<sub>1</sub> full-sibling progeny from a cross between one heterozygous plant each of ‘Grasslands Impact’ hybrid ryegrass (*L. × boucheanum* syn. *L. hybridum*) and ‘Grasslands Samson’ perennial ryegrass (*L. perenne*). Both parents are infected by New Zealand common toxic endophyte (*Epichloë festucae* var. *lolii*). Details on morphological and vegetative characteristics of the I × S population and the parents are discussed elsewhere (Sartie et al. 2009, 2011; Faville et al. 2012).

#### Field assessment of seed production traits at Palmerston North

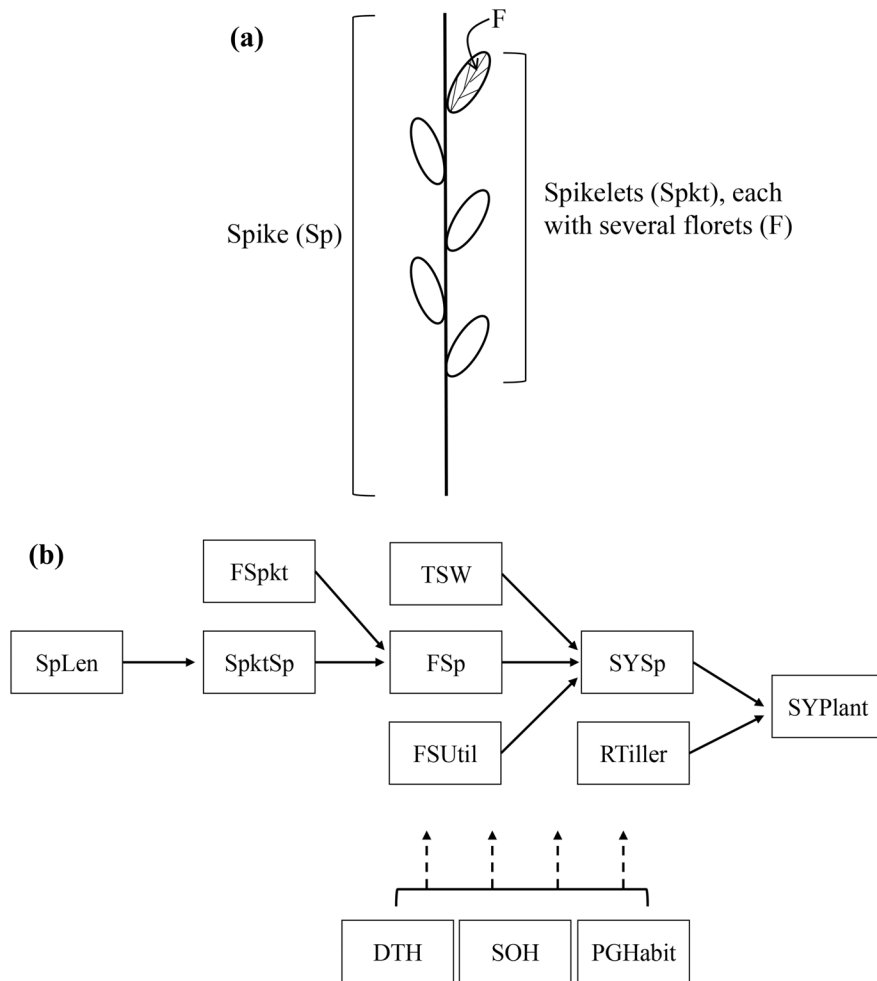
The field experiment for study of seed yield related traits (Table 1, Fig. 1) and their associations was carried out from July 2003 to February 2004 at AgResearch in Palmerston North, New Zealand (40°21’S, 175°37’E) on a fertile, alluvial silt loam soil with pH of 5.5. Four clonal replicates of the 200

mapping population progeny and of the two parents were transplanted to the field on 14–15 July 2003, using a randomised complete block design with 60 cm between plants. There was one copy of each plant genotype, comprising 10–15 tillers per replicate. A total nitrogen (N) application of 200 kg N/ha was applied as urea in three equal portions on 1st September, 1st October and 15th October. The plant growth regulator Moddus<sup>®</sup> (trinexapac-ethyl, TE) was applied at 200 g TE/ha at first spike emergence (5th November) and again three weeks later to prevent plants from lodging (Chastain et al. 2003). Stem rust (*Puccinia graminis*) was observed and controlled with Systhane 125 fungicide, applied at 20 mls/100 L of water/ha in December 2003 and January 2004.

During November and December, data were collected on heading date [DTH (2003)], spread of heading (SOH), and plant growth habit (PGHabit), as defined in Table 1. Three spikes were harvested randomly from each plant 7 weeks after heading,

**Table 1** Seed yield related traits measured, trait abbreviations, and methodology for trait assessment

Trait description	Trait abbreviation	Trait phenotype assessment
Seed yield per plant (g)	SYPlant	The weight of seed per plant measured after drying, threshing and cleaning of seeds
Seed yield per spike (mg)	SYSp	Calculated as: SYSp = (1000 × SYPlant)/RTiller
Reproductive tiller number per plant	RTiller	The total number of tillers that produced heads. Tillers with matured spikes, indicated by brown or gold coloured glumes, were counted and expressed as a percentage of total reproductive tillers
Spikelets per spike	SpktSp	The mean number of spikelets for three spikes per plant
Florets per spikelet	FSpkt	The mean number of florets on the bottom, middle and top spikelets on each of three spikes, counted post anthesis (Elgersma 1985)
1000 seed weight (g)	TSW	Measured by counting and weighing 1000 seeds, after seed was processed on a South Dakota Seed Blower to remove seed < 1.3 g TSW
Spike length (cm)	SpLen	The mean distance (cm) for three spikes per plant measured from the first culm node to the tip of the spike
Florets per spike	FSp	Calculated by multiplying values for SpktSp and FSpkt, obtained as described above
Floret site utilisation	FSUtil	The number of saleable seeds per plant divided by the number of florets per plant counted post anthesis (Elgersma 1985). Calculated as: FSUtil = (SYPlant × 1000/TSW)/(RTiller × FSp)
Days to heading from 1 November 2003	DTH (2003)	The number of days from 1 <sup>st</sup> November, to when emerging tips were visible on three spikes
Days to heading from 1 November 2004	DTH (2004)	DTH measurement was repeated on the same plants in 2004 and expressed as days after 1st November
Spread of heading (days)	SOH	The number of days between the emergence of the first and the thirtieth spike of each plant
Plant growth habit	PGHabit	Recorded after heading using a subjective scale of 1–9 (1 = erect, 3 = semi erect, 5 = medium, 7 = semi-prostrate and 9 = prostrate)



**Fig. 1** **a** Stylistic representation of the perennial ryegrass inflorescence showing key components. **b** Conceptual relationships among seed yield-related traits and seed yield per plant (SYPlant) in perennial ryegrass, showing SYPlant as a product of  $RTiller \times FSp \times FSUtil \times TSW$ .  $SYSp$  = seed yield per

spike;  $RTiller$  = no. of reproductive tillers;  $SpktSp$  = spikelets per spike,  $FSpkt$  = florets per spikelet,  $TSW$  = thousand seed weight,  $SpLen$  = spike length,  $FSp$  = florets per spike,  $FSUtil$  = floret site utilisation,  $DTH$  = days to heading,  $SOH$  = spread of heading,  $PGHabit$  = plant growth habit

placed in sealed plastic bags and kept in a refrigerator. Plants were harvested with hand-shears to about 10 cm above ground level between 6 and 26 January 2004, as seed heads ripened. In order to minimize seed shattering losses of early-emerging heads and immaturity at harvest of late-emerging heads, harvesting was carried out when most of the spikes had brown or gold coloured glumes. After harvest heads were air dried for three weeks, then stored in paper bags in a dry warehouse until they could be hand processed to determine seed-yield-related traits. The total number of reproductive tillers from each plant was counted.  $SYPlant$  and thousand seed weight ( $TSW$ ) were

measured, while data on spikelets per spike ( $SpktSp$ ),  $FSpkt$  and spike length ( $SpLen$ ) were collected from the three spikes that were harvested earlier from each plant, as defined in Table 1. Florets per spike ( $FSp$ ), seed yield per spike ( $SYSp$ ) were calculated from those data (Table 1). Floret site utilisation (the number of saleable seeds per plant divided by the number of florets counted per plant post anthesis, Elgersma 1985) was also calculated as:  $(FSUtil) = [(SYPlant, g) \times (1000/TSW, g)] / [(RTiller) \times FSp]$ . Seed moisture content was measured for four randomly selected genotypes using the high constant temperature oven

method (International Seed Testing Association, 2004) at 130 °C ( $\pm 2$  °C) for 1 h.

Plants were maintained in the field through winter 2004 and heading date measured as in the previous year [DTH (2004), Table 1] to evaluate the year-to-year repeatability of heading date differences among the genotypes.

#### Field assessment of seed production traits at Lincoln

For assessment of key seed yield traits in a second environment, population  $I \times S$  was evaluated for SYPlant, SYSp and DTH at a Lincoln site near Christchurch, New Zealand as part of an existing trial of similar design to that of the Palmerston North trial (Faville et al. 2012). Traits were evaluated as described for the Palmerston North experiment, including measurement of DTH in two consecutive years.

#### Weather data

Data on temperature, rainfall and sunshine during the period of experimentation were recorded for Palmerston North at a weather station 600 m from the experimental plots, and at Lincoln using the closest National Institute of Water and Atmospheric Research

weather station (Lincoln Broadfield, except Christchurch Aero for sunshine) (Table 2).

#### Statistical analysis

Statistical analysis of phenotypic data from each of the Palmerston North and Lincoln locations was carried out using the linear mixed models option in GenStat v 8.1 (GenStat 2005). A random linear model was applied in the analyses, using the REML algorithm, with  $I \times S$  genotype, replicate, row and column effects considered as random. Adjusted phenotypic means were based on Best Linear Unbiased Predictors (BLUP's) (White and Hodge 1989), enabling adjustment for random error across replicates, columns and rows within replicates. The significance of an estimated variance component was determined by the ratio of the component relative to its standard error. If the variance component of a model term was more than two standard errors from zero, then the variance component was considered significant ( $p < 0.05$ ). After inspection of normal probability plots, data transformation was not considered necessary.

The Palmerston North dataset contained 1616 observations for DTH (200  $F_1$  progeny + 2 parents = 202 genotypes  $\times$  2 years  $\times$  4 replicates) and 808 observations for each of the other eleven traits which were measured in 1 year only. Similarly, the Lincoln dataset consisted of 1616 observations for

**Table 2** Mean monthly temperature, rainfall and sunshine observed in close proximity to the Palmerston North (PN) and Lincoln (LIN) experimental sites during July 2003 to January 2004

Month	Air temp (°C) (max) <sup>a</sup>		Air temp (°C) (min) <sup>a</sup>		Soil temp (°C) (10 cm)		Rainfall (mm/month)		Sun (h/day)	
	PN	LIN	PN	LIN	PN	LIN	PN	LIN	PN	LIN
July (2003)	12.7	10.9	1.6	0.3	5.9	2.9	83.7	53.2	5.5	4.5
August (2003)	14.8	11.9	4.7	2.5	7.9	5.0	27.9	48.4	5.5	4.7
September (2003)	15.3	14.5	6.3	4.6	10.3	7.0	15.3	88.2	4.3	5.5
October (2003)	16.4	16.0	6.2	5.6	–	9.1	89.9	27.2	–	7.6
November (2003)	17.1	18.7	8.8	7.2	13.2	12.8	81.0	35.6	4.3	9.1
December (2003)	20.3	21.8	11.6	10.4	16.1	16.9	80.6	1.2	4.4	8.9
January (2004)	20.5	23.4	8.4	13.3	15.8	19.0	27.9	21.2	6.9	7.0
Mean	16.7	16.7	6.8	6.3	11.5	10.4	58.0	39.3	5.2	6.7

Data for soil temperature and sunshine were not available for October 2003 at PN and were not included in the calculation of the mean values. Sunshine data for LIN were acquired from a station more distant from the experimental site than that used for other parameters

<sup>a</sup>Mean daily maxima and minima

DTH and 808 observations for each of SYPlant, SYSp and RTiller. Broad sense heritability for each trait was calculated from variance components yielded by the random effects model as:

$$H_b = \frac{\sigma_g^2}{\sigma_g^2 + \left(\frac{\sigma_e^2}{n_r}\right)}$$

where  $\sigma_g^2$  = genotypic component of variance,  $\sigma_e^2$  = residual variance of genotypes and  $n_r$  = number of replications (Burton and DeVane 1953). Trait associations were evaluated by correlation and principal component analysis (PCA) (Sartie et al. 2011) using Minitab version 10.51 (Minitab Inc, 2081 Enterprise Drive, State College, PA).

Analysis of a dataset combining trait data from the Palmerston North and Lincoln sites for DTH, SYPlant, SYSp and RTiller was also completed, using the linear mixed models option in GenStat as described above and with site additionally considered as a fixed effect. The linear model also included a genotype-by-site interaction effect. Heritability was estimated from variance components as:

$$H_b = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gs}^2}{n_s} + \left(\frac{\sigma_e^2}{n_r n_s}\right)}$$

where  $\sigma_{gs}^2$  = genotype-by-site component of variance and  $n_s$  = number of sites.

### QTL analysis

Quantitative trait loci analysis was conducted using the I × S consensus genetic linkage map marker data (Sartie et al. 2011). The mean value for each trait from 188 F<sub>1</sub> genotypes in the present study was used for QTL analysis implemented in MapQTL<sup>®</sup> 4.0 software (Van Ooijen et al. 2002). Simple interval mapping (IM) was performed and then multiple QTL model (MQM) mapping was used to refine the position and magnitude of QTL. The default mapping step size of 5 cM was used for both IM and MQM. Co-factors for MQM analysis were selected using a procedure based on forward selection followed by backward elimination (Van Ooijen et al. 2002) as described in Sartie et al. (2011).

Permutation testing (n = 1000) was performed for each trait to establish LOD thresholds for QTL declaration at a linkage group- or genome-wide

significance at  $p < 0.05$  (Churchill and Doerge 1994). QTL position was described by LOD peak position and 1- and 2-LOD support intervals. An additional criterion for declaration of a significant QTL was the presence of markers within the 2-LOD support interval that were significantly associated with the trait in a non-parametric Kruskal–Wallis analysis, executed in MapQTL<sup>®</sup> 4.0. QTL names are constructed as: trait name-year of trait assessment-linkage group (LG) number.

As described previously (Sartie et al. 2011), phenotypic trait means for the four different QTL genotype classes at a locus (ac, ad, bc, and bd) calculated in MapQTL 4.0 were used to report QTL in terms of the individual parental effects (i.e. the difference in effect of the alleles inherited from each parent, ‘I’ and ‘S’), following the model of Knott et al. (1997) as used by Sewell et al. (2000, 2002).

For mapped ryegrass ESTs, sequence alignment by the Basic Local Alignment Search Tool, BLASTN (Altschul et al. 1990, 1997) (threshold values of < E-15; SID > 85% over > 100 bp) was used to estimate identity with homologous rice genome positions in the MSU rice pseudomolecule assembly hosted at [www.gramene.org](http://www.gramene.org). Where possible the MSU assembly was also used to identify rice genome positions for QTL-linked markers reported from previous studies.

## Results

### Phenotype analysis

#### Palmerston North experiment

Data are presented in a logical rather than chronological collection sequence: SYPlant, SYSp as the overarching traits, measured primary data (RTiller, SpktSp, FSpkt, TSW and SpLen), derived secondary data (FSp, FSUtil), and other measures (DTH, SOH, PGHabit). A majority of traits exhibited statistically significant differences between the parents and among the progeny, and transgressive segregation (Table 3). The range among genotypes in SYPlant was greater than 5×, whereas for the primary yield component traits the range was smaller (RTiller 3.2×, SpktSp 1.5×, FSpkt 1.6×). Although parent ‘I’ (from ‘Grasslands Impact’) had more tillers per plant pre-flowering than parent ‘S’ (‘Grasslands Samson’) (data not



**Table 3** Mean ( $\pm$  SE, standard error of the mean), range, genotype variance component ( $\sigma_g^2 \pm$  SE), genotype-by-environment variance component ( $\sigma_{gs}^2 \pm$  SE) and broad sense heritability ( $H_b$ ) for seed yield-related traits of the I  $\times$  S perennial ryegrass mapping population progeny and parentsassessed in the field as spaced plants in 2003 and 2004, at Palmerston North and Lincoln. S: 'Samson' parent; I: 'Impact' parent.  $LSD_{0.05}$  = least significant difference amongst I  $\times$  S genotypes at  $p < 0.05$ 

Trait <sup>a</sup>	Mean	Range	S	I	$LSD_{0.05}$	$\sigma_g^2$	$\sigma_{gs}^2$	$H_b$
<i>Palmerston North</i>								
SYPlant (g)	35.3 ( $\pm$ 5.11)	11.9–64.5	27.7	11.9	14.47	77.9 $\pm$ 10.5	–	0.75
SYSp (mg)	91.9 ( $\pm$ 10.85)	46.7–169.6	103.6	65.1	30.61	358.9 $\pm$ 48.2	–	0.75
RTiller	391.6 ( $\pm$ 51.04)	184.0–592.0	272.0	184.0	144.90	3922 $\pm$ 674	–	0.59
SpktSp	26.7 ( $\pm$ 1.36)	21.0–31.0	29.0	23.0	3.81	0.7 $\pm$ 0.3	–	0.27
FSpkt	8.4 ( $\pm$ 0.59)	7.0–11.0	8.0	7.0	1.69	0.17 $\pm$ 0.06	–	0.33
TSW (mg)	1.9 ( $\pm$ 0.07)	1.4–2.5	2.1	1.4	0.27	0.03 $\pm$ 0.00	–	0.76
SpL (cm)	19.6 ( $\pm$ 0.95)	16.1–23.6	19.0	19.8	2.68	0.6 $\pm$ 0.2	–	0.39
FSp	225.8 ( $\pm$ 21.56)	162.0–298.0	224.0	162.0	61.17	230 $\pm$ 75	–	0.33
FSUtil	0.2 ( $\pm$ 0.04)	0.1–0.4	0.3	0.3	0.03	0.0017 $\pm$ 0.0003	–	0.94
DTH (2003) (d)	26.0 ( $\pm$ 0.93)	13.0–36.0	18.0	36.0	2.64	12.9 $\pm$ 1.4	–	0.94
DTH (2004) (d)	20.0 ( $\pm$ 0.95)	11.0–30.0	18.0	27.0	1.02	0.4 $\pm$ 0.1	–	0.77
SOH (d)	6.6 ( $\pm$ 1.08)	2.0–13.0	13.0	7.0	3.06	1.6 $\pm$ 0.3	–	0.58
PGHabit	3.3 ( $\pm$ 0.49)	1.0–7.0	1.0	7.0	1.39	0.8 $\pm$ 0.1	–	0.77
<i>Lincoln</i>								
SYPlant (g)	24.0 ( $\pm$ 4.76)	15.6–36.1	23.1	26.4	9.52	37.8 $\pm$ 8.1	–	0.44
SYSp (mg)	100.6 ( $\pm$ 14.36)	77.6–187.1	101.2	101.3	28.72	322.0 $\pm$ 77.0	–	0.40
RTiller	64.2 ( $\pm$ 10.02)	43.1–161.2	60.4	65.9	20.04	188.5 $\pm$ 34.2	–	0.48
DTH (2003) (d)	16.4 ( $\pm$ 1.36)	9.9–23.8	13.4	19.7	2.73	7.6 $\pm$ 0.8	–	0.77
DTH (2004) (d)	7.8 ( $\pm$ 1.18)	4.5–11.2	6.8	9.0	2.35	2.7 $\pm$ 0.5	–	0.50
<i>Across environments</i>								
SYPlant (g)	29.5 ( $\pm$ 5.63)	15.4–48.2	24.8	22.8	10.72	31.4 $\pm$ 7.6	29.5 $\pm$ 7.3	0.48
SYSp (mg)	96.4 ( $\pm$ 13.43)	67.4–135.5	101.8	85.6	26.84	239.0 $\pm$ 44.0	71.0 $\pm$ 35.0	0.60
DTH (2003) (d)	19.3 ( $\pm$ 2.79)	11.7–26.3	13.4	27.0	2.78	7.8 $\pm$ 1.0	2.0 $\pm$ 0.4	0.79
DTH (2004) (d)	13.6 ( $\pm$ 1.56)	9.1–18.7	12.3	16.3	2.56	2.8 $\pm$ 0.5	1.0 $\pm$ 0.3	0.64

<sup>a</sup>SYPlant = seed yield per plant; SYSp = seed yield per spike; RTiller = no. of reproductive tillers; SpktSp = spikelets per spike; FSpkt = florets per spikelet; TSW = thousand seed weight; SpLen = spike length; FSp = florets per spike; FSUtil = floret site utilisation; DTH = days to heading; SOH = spread of heading; PGHabit = plant growth habit

shown), the latter had more reproductive tillers (RTiller) and greater seed yield per plant (SYPlant) than parent 'I' (Table 3).

When trait associations are assessed by simple correlation analysis (Table 4) SYPlant is most strongly associated with seed yield per spike (SYSp) and RTiller. SYSp in turn is most closely correlated with floret site utilisation (FSUtil) while RTiller displays a negative correlation with spread of heading (SOH). The only significant negative correlations between seed yield components involve FSUtil, and these were modest, but greater SOH is associated with

lower SYPlant, later heading with lower florets per spikelet (FSpkt), and prostrate growth habit (PGHabit) with lower TSW (Table 4).

When trait associations are assessed by principal component analysis to extract trait associations corrected for the effects of variables on each other (Table 5), the first three principal components (PCs) collectively account for 57.9% of the data variation and show a progressively diminishing link with SYPlant (coefficients 0.479, 0.257, and 0.172, respectively) driven largely by FSpkt contribution to FSp and SYSp (PC1), retention of FSUtil in competition with

**Table 4** Coefficients of correlation between seed yield component traits of the I × S perennial ryegrass mapping population assessed in the field as spaced plants at Palmerston North and Lincoln (suffix ‘Lin’); ( $p < 0.05$  if  $r \geq 0.14$ ;  $p < 0.01$  if  $r \geq 0.18$ ; ns = non-significant,  $p > 0.05$ )

	SYPlant <sup>a</sup>	SYSp	RTiller	SpktSp	FSpkt	TSW	SpLen	FSp	FSUtil	DTH 03	DTH 04	SOH	PGHabit	SYPlant Lin	SYSp Lin	RTiller Lin	DTH Lin
SYSp	0.738																
RTiller	0.617	ns															
SpktSp	ns	0.233	ns														
FSpkt	0.232	0.151	0.219	0.268													
TSW	0.190	0.237	ns	ns	ns												
SpLen	0.154	0.233	ns	0.156	ns	ns											
FSp	0.239	0.220	0.149	0.693	0.815	ns	0.149										
FSUtil	0.471	0.743	-0.143	ns	-0.173	ns	ns	-0.181									
DTH 03	ns	ns	ns	-0.162	-0.408	ns	ns	-0.430	0.212								
DTH 04	ns	ns	ns	ns	-0.319	ns	ns	-0.242	0.256	0.663							
SOH	-0.453	-0.259	-0.365	ns	ns	ns	ns	ns	-0.285	-0.427							
PGHabit	ns	ns	ns	ns	ns	-0.185	ns	ns	ns	ns	ns	-0.199					
SYPlant Lin	0.302	0.323	ns	ns	ns	ns	ns	ns	0.315	ns	ns	-0.145	ns				
SYSp Lin	0.303	0.424	ns	ns	ns	0.158	ns	ns	0.308	ns	ns	ns	ns	0.707			
RTiller Lin	0.194	0.184	ns	ns	ns	ns	ns	ns	ns	ns	ns	-0.203	ns	0.246	0.184		
DTH Lin 03	ns	ns	ns	-0.168	-0.331	ns	ns	-0.329	0.150	0.649	0.507	-0.242	ns	ns	ns	ns	
DTH Lin 04	0.157	ns	ns	ns	-0.183	ns	ns	-0.207	0.210	0.543	0.496	-0.329	ns	0.380	0.166	0.197	0.365

<sup>a</sup>SYPlant = seed yield per plant; SYSp = seed yield per spike; RTiller = no. of reproductive tillers; SpktSp = spikelets per spike; FSpkt = florets per spikelet; TSW = thousand seed weight; SpLen = spike length; FSp = florets per spike; FSUtil = floret site utilisation; DTH = days to heading; SOH = spread of heading; PGHabit = plant growth habit



**Table 5** Coefficients of principal component analysis of seed yield per plant and component traits for the I × S perennial ryegrass mapping population assessed in the field at Palmerston North as spaced plants in 2003. Coefficients of absolute value less than 0.15 have been suppressed

Component	PC1 (24.5%)	PC2 (20.9%)	PC3 (12.5%)	PC4 (10.8)
SYPlant <sup>a</sup>	0.479	0.257	0.172	0.189
SYSp	0.424	0.289	– 0.352	–
RTiller	0.242	–	0.655	0.212
SpktSp	0.294	– 0.246	– 0.225	– 0.160
FSpkt	0.349	– 0.361	–	–
TSW	–	–	–	0.707
SpLen	0.176	–	– 0.301	0.180
FSp	0.404	– 0.408	–	–
FSUtil	0.212	0.439	– 0.323	– 0.225
DTH (2003)	0.167	0.392	–	–
SOH	– 0.204	– 0.359	– 0.352	0.152
PGHabit	–	–	0.153	– 0.517

<sup>a</sup>SYPlant = seed yield per plant; SYSp = seed yield per spike; RTiller = no. of reproductive tillers; SpktSp = spikelets per spike; FSpkt = florets per spikelet; TSW = thousand seed weight; SpLen = spike length; FSp = florets per spike; FSUtil = floret site utilisation; DTH = days to heading; SOH = spread of heading; PGHabit = plant growth habit

spikelets per spike (SpktSp) and FSpkt (PC2), and high RTiller (PC3). PC4 (10.8% of data variation) is of interest as it indicates only a modest link between 1000-seed weight (TSW) and SYPlant (Elgersma 1990b), but confirmation of a positive relationship between high TSW and erect growth habit (Table 5).

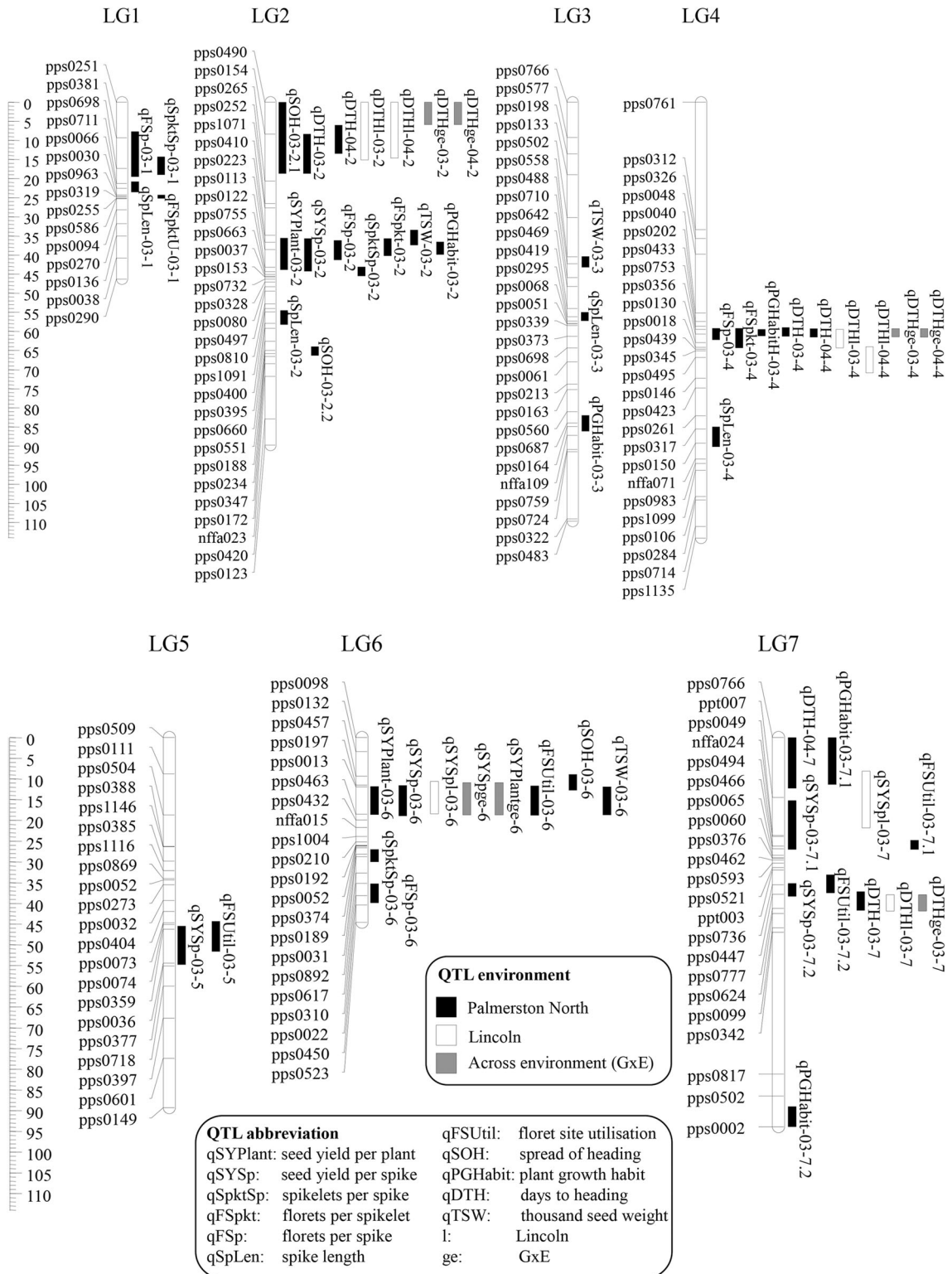
#### Lincoln experiment and across-site analysis

Data for SYPlant, SYSp, RTiller and days to heading (DTH) from Lincoln also exhibited significant genotypic variation (Table 3). Mean trait values and ranges amongst IxS genotypes for Lincoln data was similar to that observed in Palmerston North, except that the range for SYPlant was smaller and mean RTiller was considerably lower than in Palmerston North. SYPlant, SYSp, DTH (2003) and DTH (2004) from Lincoln significantly ( $p < 0.05$ ) correlated with their respective Palmerston North data (Table 4), with the highest correlations involving DTH ( $r = 0.50$ – $0.65$ ). Of the seed yield traits, the highest correlation between sites was for SYSp ( $r = 0.43$ ). RTiller from Lincoln did not correlate significantly with Palmerston North RTiller. Within the Lincoln site, the magnitude of correlations amongst traits was similar to that observed for Palmerston North but RTiller was only weakly correlated with SYPlant at Lincoln.

Across-site statistical analysis resulted in significant  $\sigma_{gs}^2$  components (Table 3) for the four measured traits, confirming a  $G \times E$  influence for all. Based on the ratio of  $\sigma_{gs}^2/\sigma_g^2$ , the  $G \times E$  effect was strongest for SYPlant. SYPlant had an interaction ratio of 0.94 compared with 0.29, 0.25 and 0.36 for SYSp, DTH (2003) and DTH (2004), respectively. This was broadly consistent with the estimated heritabilities for these traits (Table 3).

#### QTL analysis

In total 42 significant QTL, located at 21 discrete genomic positions, were detected by MQM analysis using data from the Palmerston North experiment (Fig. 2). QTL were identified for all the traits except RTiller and were present on all seven linkage groups (LG), with two to five significant QTL declared per trait. QTL were generally of low or moderate effect, explaining on average 12.3% of total phenotypic variation ( $V_p$ ). Major QTL were identified for some traits, notably DTH for which a QTL accounting for 39% (2003) and 28% (2004)  $V_p$  was detected on LG2 (Table 6). Total  $V_p$  explained by all QTL for a trait ranged from 19.8% (FSpkt) to 74.9% (PGHabit), with the mean across all traits 40.2%. However, the true magnitudes of QTL effects in this study are likely to be



◀ **Fig. 2** Genetic linkage map developed for  $F_1$  perennial ryegrass mapping population  $I \times S$ , showing QTL for seed yield and seed yield related traits, as measured in the field using spaced plants at Palmerston North, New Zealand in 2003 and 2004. QTL for selected traits at the Lincoln site and across environments are also shown. Marker names are shown at left of linkage groups (LG) and QTL locations (2-LOD confidence intervals) are indicated by rectangles at right of each LG (QTL names as per Table 6). The length of linkage groups in centimorgans (cM) is indicated by the scale at the left of the figure

smaller than reported as the analysis was conducted in a relatively small population ( $n = 188$ ) (Elgersma 1990a), making  $V_p$  values prone to upwards bias (Beavis 1994).

Eleven of the 42 QTL occurred in positions independent of other traits, most notably SpLen for which all four QTL occurred at positions free of QTL for other traits (Fig. 2). By contrast, clustering of multiple QTL (2-LOD interval overlap) occurred at ten genome locations (Fig. 2), most notably on LG2, 4 and 6, indicating a potential common genetic basis for QTL at these positions. QTL for SYPlant on LG2 and 6 both co-located with QTL for SYSp and other component or related traits.

The LG2 QTL cluster identifies relationships amongst SYPlant with SYSp, FSpkt, SpktSp, FSp, TSW and PGHabit. Estimated parental effects (phase and size) were conserved between SYPlant and SYSp (Table 6), with a strong effect due to alleles segregating from parent 'I'. Consistent parental effects were also estimated for FSp, FSpkt and SpktSp and PGHabit at this position, but these differed from SYPlant/SYSp in that approximately equal effects were conferred by alleles segregating from both parents. For TSW the alleles segregating from parent 'S' exerted a large effect, opposite to that estimated for the other traits. Differing patterns of parental effects amongst QTL at the LG2 position may be due to the influence of alternative alleles at the QTL or it is possible that this reflects tight linkage amongst separate loci modulating the expression of the separate traits. It is not possible, with the resolution of QTL mapping achieved in this study, to decisively determine which of these scenarios holds.

The LG6 position identified a relationship of TSW, FSUtil and SYSp with SYPlant. Here the parental effects were consistent amongst all QTL (Table 6),

implying a common genetic basis. A QTL for SOH was detected in close proximity (Fig. 2) and may be a product of the same genetic locus, but was characterised by an opposing set of parental effects (Table 6). High SOH is likely to lead to high seed loss before harvest and thus links to low FSUtil and low seed yield.

Large effect QTL for DTH on LG2 and LG4 were detected in both years (2003 and 2004), while smaller effect QTL were detected at different positions on LG7 in 2003 and 2004, respectively (Fig. 2). DTH (2003) and DTH (2004) QTL at the LG2 and LG4 positions shared consistent additive parental effects (Table 6). QTL for other traits at these positions had parental effects in opposition to those of DTH (SOH on LG2; FSpkt and FSp on LG4), in line with the negative correlations observed between these traits and DTH (Table 4).

The strong correlation observed between FSUtil and SYSp (Table 4) was manifested in the co-alignment of QTL for these traits at four positions on LG5, 6 and 7. In all positions, QTL parental effects were conserved for the two traits, suggesting a common genetic factor.

QTL analysis of data for the four traits measured at Lincoln, revealed a total of seven QTL, for SYSp and DTH (Fig. 2, Table 6). No QTL were detected for RTiller, as for Palmerston North, and additionally no significant QTL for Lincoln SYPlant were found. The QTL for SYSp (LG6 and LG7), DTH 2003 (LG2, LG4, LG7) and DTH 2004 (LG2 and LG4) (Table 6) co-located with their Palmerston North equivalents; no Lincoln-specific QTL were detected for any trait. In all cases these additional QTL shared similar direction and magnitude of parental effects with their counterparts discovered in Palmerston North data, suggesting a common genetic basis (Table 6). QTL analysis based on BLUP means from a GxE analysis for SYPlant, SYSp and DTH confirmed QTL positions for all three traits identified by the Palmerston and Lincoln analyses (Fig. 1, Table 6).

## Discussion

Associations between seed yield component traits

Seed yield is a complex trait, and the identification of its component traits and their interactions, and related

**Table 6** QTL controlling seed yield and related traits, detected on a genetic linkage map by simple interval mapping (IM) and multiple QTL model (MQM) analyses in I × S F<sub>1</sub> perennial ryegrass mapping population data in 2003 and 2004 at Palmerston North, New Zealand

Trait	QTL	LG	LOD threshold		Peak LOD score		V <sub>p</sub> (%)	LOD peak (cM ± 2LOD)	Closest marker	Parental effects	
			LG wide	Genome wide	IM	MQM				I	S
SYPlant	qSYPlant-03-2	2	2.75	3.73	3.88	5.45	10.2	38.6 (35.6–43.9)	pps0113	12.7	– 0.6
	qSYPlant-03-6	6	2.52	3.73	6.67	6.89	13.1	16.9 (11.8–18.6)	pps0463	13.0	– 7.9
SYSp	qSYSp-03-2	2	2.92	3.91	3.57	4.36	7.9	38.6 (35.7–44.2)	pps0113	21.0	– 1.4
	qSYSp-03-5	5	2.85	3.91	4.25	4.98	9.5	54.4 (45.5–54.8)	pps0377	17.0	– 10.9
	qSYSp-03-6	6	2.70	3.91	10.25	11.73	18.6	16.9 (11.6–18.9)	pps0463	33.8	– 19.7
	qSYSp-03-7.1	7	2.95	3.91	< 2.00	4.04	9.8	26.8 (15.2–27.0)	pps0466	– 49.8	0.9
SpktSp	qSYSp-03-7.2	7	2.95	3.91	< 2.00	4.42	10.9	37.8 (35.2–38.3)	pps0447	53.0	– 0.2
	qSpktSp-03-1	1	2.58	3.70	3.95	5.25	11.5	17.3 (14.3–19.0)	pps0698	– 1.19	– 1.29
	qSpktSp-03-2	2	2.90	3.70	2.78	3.87	8.0	45.1 (43.1–45.5)	pps0663	0.85	1.52
	qSpktSp-03-6	6	2.67	3.70	3.75	6.50	15.5	28.7 (27.0–30.0)	pps0892	1.90	– 0.23
	qFSpkt-03-2	2	2.92	3.84	3.61	4.21	9.1	38.6 (35.7–40.2)	pps0113	0.71	0.52
TSW	qFSpkt-03-4	4	3.00	3.84	4.12	4.72	10.7	60.7 (59.2–64.3)	pps0356	– 0.75	– 0.69
	qTSW-03-2	2	3.00	3.88	4.25	5.14	10.5	36.7 (33.5–37.4)	pps0223	– 0.05	– 0.22
	qTSW-03-3	3	2.93	3.88	7.20	8.81	19.2	42.3 (40.4–43.2)	pps0488	0.04	– 0.30
	qTSW-03-6	6	2.62	3.88	3.93	4.69	8.9	18.5 (11.9–18.7)	pps0463	0.13	– 0.18
SpLen	qSpLen-03-1	1	2.44	3.77	3.31	4.18	5.2	22.5 (20.8–23.5)	pps0066	– 0.33	0.78
	qSpLen-03-2	2	2.84	3.77	2.77	3.92	5.7	57.9 (54.5–58.2)	pps0395	– 1.20	– 0.17
	qSpLen-03-3	3	2.90	3.77	8.99	13.00	18.8	56.2 (55.0–57.2)	pps0419	2.31	0.13
FSp	qSpLen-03-4	4	2.92	3.77	3.17	3.88	6.4	89.3 (85.0–90.2)	pps0150	1.19	– 0.48
	qFSp-03-1	1	2.38	3.82	3.97	4.61	9.9	17.3 (7.7–19.5)	pps0698	– 6.3	– 26.6
	qFSp-03-2	2	2.82	3.82	3.92	4.95	9.0	38.6 (36.2–41.3)	pps0113	21.4	20.4
	qFSp-03-4	4	2.80	3.82	3.45	4.26	7.7	60.7 (59.2–62.2)	pps0356	– 26.6	– 16.2
FSUtl	qFSp-03-6	6	2.64	3.82	2.47	3.88	7.3	38.1 (35.3–39.9)	pps0022	28.1	– 4.8
	qFSUtl-03-1	1	2.40	3.88	3.36	4.01	6.6	24.9 (24.3–25.1)	pps0319	– 0.041	0.036
	qFSUtl-03-5	5	2.74	3.88	4.17	5.59	9.8	46.2 (44.9–51.6)	pps0359	0.012	– 0.025
	qFSUtl-03-6	6	2.77	3.88	5.23	5.91	9.0	11.5 (11.7–18.7)	pps0013	0.045	– 0.053
	qFSUtl-03-7.1	7	2.68	3.88	< 2.00	4.39	8.8	26.2 (24.8–27.0)	pps0466	– 0.133	0.018
DTH (2003)	qFSUtl-03-7.2	7	2.68	3.88	< 2.00	4.09	8.3	35.5 (33.1–37.5)	pps0736	0.126	0.007
	qDTH-03-2	2	2.90	3.78	11.69	21.20	38.9	13.4 (8.4–18.7)	pps0154	2.27	– 9.23
	qDTH-03-4	4	3.00	3.78	6.94	16.45	16.8	60.7 (59.0–61.3)	pps0356	3.02	6.03
	qDTH-03-7	7	2.73	3.78	3.18	6.69	6.2	41.2 (37.2–41.7)	pps0777	– 0.32	– 3.69

**Table 6** continued

Trait	QTL	LG	LOD threshold		Peak LOD score		V <sub>p</sub> (%)	LOD peak (cM ± 2LOD)	Closest marker	Parental effects	
			LG wide	Genome wide	IM	MQM				I	S
DTH (2004)	qDTH-04-2	2	2.76	3.72	9.37	13.49	28.4	8.4 (6.0–13.4)	pps0154	0.12	- 4.96
	qDTH-04-4	4	2.88	3.72	3.70	7.97	13.7	60.5 (59.3–61.4)	pps0356	1.96	2.66
	qDTH-04-7	7	2.64	3.72	2.80	4.44	7.3	0.0 (0.0–12.2)	pps0766	- 1.21	- 1.99
	qSOH-03-2.1	2	2.78	3.77	3.10	4.96	10.0	8.4 (0.0–18.7)	pps0154	- 0.77	2.03
	qSOH-03-2.2	2	2.78	3.77	3.63	6.92	15.7	65.8 (64.0–66.3)	pps0234	- 1.52	- 2.02
	qSOH-03-6	6	2.68	3.77	2.63	4.28	5.6	11.5 (8.9–11.6)	pps0197	- 1.72	0.24
PGHabit	qPGHabit-03-2	2	3.03	3.85	5.58	11.07	15.8	38.6 (36.7–39.5)	pps0113	1.17	1.30
	qPGHabit-03-3	3	2.98	3.85	2.61	4.30	7.3	84.0 (82.0–86.1)	pps0687	- 0.88	- 0.75
	qPGHabit-03-4	4	2.97	3.85	5.99	13.03	23.1	60.7 (59.5–61.2)	pps0356	1.28	- 1.05
	qPGHabit-03-7.1	7	2.74	3.85	2.66	8.91	18.1	0.0 (0.0–11.3)	pps0766	0.83	1.25
	qPGHabit-03-7.2	7	2.74	3.85	2.72	4.37	10.6	93.9 (89.0–93.9)	pps0002	0.77	0.09
	qDTHI-03-2	2	2.81	3.75	3.96	6.42	12.6	0.0 (0.0–14.6)	pps0490	3.07	- 3.19
DTH (2003 Lin)	qDTHI-03-4	4	2.89	3.75	8.67	10.54	17.6	62.9 (59.6–64.2)	pps0130	4.32	3.50
	qDTHI-03-7	7	2.79	3.75	2.65	4.08	2.7	41.2 (37.0–41.9)	pps0777	0.31	- 2.54
	qDTHI-04-2	2	2.89	3.81	9.49	10.76	23.7	0.0 (0.0–14.3)	pps0490	3.07	- 3.90
	qDTHI-04-4	4	2.91	3.81	3.49	5.38	10.1	65.1 (64.0–71.8)	pps0345	2.13	1.83
	qSYSpl-04-6	6	2.78	3.95	4.67	5.10	12.2	11.9 (10.4–18.1)	pps0013	31.9	- 11.00
	qSYSpl-04-7	7	2.83	3.95	2.75	4.44	13.8	14.4 (8.0–21.5)	ppt007	- 11.70	12.24
DTH (2003 GxE)	qDTHge-03-2	2	2.88	3.82	9.40	15.40	28.1	5.0 (0.0–5.5)	pps0490	1.32	- 5.75
	qDTHge-03-4	4	2.98	3.82	9.12	13.28	23.4	60.7 (59.4–61.2)	pps0356	3.26	4.03
	qDTHge-03-7	7	2.78	3.82	3.33	5.30	5.2	41.2 (37.7–41.7)	pps0777	0.13	- 3.03
	qDTHge-04-2	2	2.84	3.80	12.50	16.22	30.7	5.0 (0.0–5.4)	pps0490	1.11	- 3.25
	qDTHge-04-4	4	2.94	3.80	3.71	6.76	10.7	60.7 (59.2–61.4)	pps0356	1.47	1.56
	qSYSpge-6	6	2.78	3.90	9.08	9.93	19.9	16.9 (11.8–18.7)	pps0463	18.67	- 14.28
SYPlant (GxE)	qSYPlantge-6	6	2.58	3.80	5.23	6.89	13.7	16.9 (11.8–18.8)	pps0463	7.08	- 6.11

QTL for selected traits detected in data from other sources (Lin = Lincoln, New Zealand; G × E = from combined sites analysis) are also provided at the bottom of the table. QTL naming convention is described in the main text; for Lincoln ('1') and G × E (ge) QTL an additional identifier is added after the trait name. LG = linkage group; LOD threshold = logarithm of the odds score for declaring significant QTL at  $p \leq 0.05$ ; V<sub>p</sub> = phenotypic variation explained by QTL; Parental effects: I = substitution effect of alleles from parent '1'; S = substitution effect of alleles from parent 'S'; sign indicates direction of effect, number indicates magnitude of effect

\*SYPlant = seed yield per plant; SYSp = seed yield per spike; SpktsSp = spikelets per spike; FSpkt = florets per spikelet; TSW = thousand seed weight; SpLen = spike length; FSp = florets per spike; FSUtil = floret site utilisation; DTH = days to heading; SOH = spread of heading; PGHabit = plant growth habit

traits and the genetic factors controlling them, will provide a better understanding for the genetic improvement of seed yield potential in perennial ryegrass. Conceptually, seed yield per plant (SYPlant) can be understood as the product of number of reproductive tillers (RTiller) and seed yield per spike (SYSp). SYSp in turn can be understood as the product of florets per spike (FSp), floret site utilisation (FSUtil) and 1000-seed weight (TSW) (Fig. 1b). FSp is the product of florets per spikelet (FSpkt) and spikelets per spike (SpktSp). Spike length (SpLen) may also influence SpktSp, while there is no logical basis for predicting the influence of DTH, SOH and PGHabit on these other traits. The statistics presented here for these traits are not all derived independently, with some being arithmetic functions of others.

By simple correlation analysis SYPlant is most closely correlated to SYSp (Table 4), but this is as much a mathematical as a biological association because SYSp equals SYPlant/RTiller. Using PCA to isolate trait associations contributing independently to SYPlant is revealing. Firstly, it is seen that the associations among the component traits of seed yield are comparatively complex. There is not a “single size” PC encompassing most of the traits and which describes plants with either superior or inferior seed yield (as often occurs in PCA analyses of this kind). Rather, a series of four components was identified (Table 5), each describing an association between a specific subset of seed yield component traits, and each explaining a part of the observed variation for seed yield amongst genotypes. PC1 is rather strongly associated with SYPlant (coefficient + 0.479) and with three of the four measured primary seed yield component traits; RTiller, SpktSp, and especially FSpkt (coefficient + 0.349). PC2 can be seen as identifying intra-plant competition factors, such as a resource-allocation trade-off reflected by a negative association between increased FSUtil (coefficient + 0.439) and decreased SpktSp and FSpkt (coefficients - 0.246 and - 0.361, respectively). This indicates that plants forming fewer florets and attaining higher FSUtil tended to have higher SYPlant (coefficient + 0.257). This finding corroborates research of Rumball and Foote (2008), where it was found that selection for inflorescence branching to increase FSp reduced SYPlant. A recent study (Abel et al. 2017), based on path and correlation analysis of seed yield components also identified SpktSp and

FSpkt as traits with a direct effect on seed yield in perennial ryegrass.

PC3 involves SOH (coefficient - 0.352), a trait not generally considered in discussion of seed yield potential. This indicates that plants with greater uniformity in flowering date (i.e. less spread) generally produce more reproductive tillers (coefficient + 0.655) with a modest contribution to SYPlant (coefficient + 0.172). Finally, PC4 is of high biological interest because it is the first to substantively involve TSW (coefficient + 0.707), a trait widely held to correlate with seed vigour, and anecdotally presumed to be important to seed yield. However, this PC accounts for < 11% of the data variation and only a modest contribution to SYPlant (coefficient + 0.189) is indicated. Indeed, TSW is more strongly associated in PC4 with erect growth habit (PGHabit coefficient - 0.517), than with SYPlant (Table 5). These findings indicate that TSW is not an important determinant of seed yield in perennial ryegrass. Other reports (Rowarth et al. 1999) show that for forage species, TSW is a seed quality factor associated with increased seedling weight (Jin et al. 1996), vigour (Bean 1980) and field emergence (Rowarth and Sanders 1996), and hence improved field production (Hampton 1986), making it a useful target for breeders.

Previous studies by several other researchers have identified SYSp and RTiller as traits that significantly influence seed yield in perennial ryegrass (Bugge 1987; White 1990; Elgersma 1990b; Marshall and Wilkins 2003; Studer et al. 2008). However, the recognition that SYSp is itself a complex trait, determined by primary traits FSUtil, SpktSp, FSpkt and TSW, and that there are several associations amongst subsets of component traits contributing independently to SYPlant, improves our understanding of seed yield in perennial ryegrass.

#### Indications for breeding targets

From a perspective of prioritising breeding targets aimed at improving seed yield in perennial ryegrass, the present data provide a number of firm indications.

Firstly, in deconstructing SYPlant and SYSp to more fundamental components, an obvious trait of interest is FSUtil. The trait has high heritability (0.94, Table 3), a high correlation with SYPlant (Table 4) and contributes to SYPlant via independent trait associations in PC1 and PC2 (Table 5). However,

there is a practical problem in that determining FSUtil in large numbers of plants for selection purposes would impose severe logistical difficulties. The possible use of QTL markers to indirectly select for FSUtil is discussed below. Another possible avenue for FSUtil enhancement is indirectly, by leveraging the modest association observed between FSUtil and DTH (i.e. higher FSUtil associated with later heading: Table 4; PC2 in Table 5). DTH also has high heritability (0.94, Table 3) and selecting for later heading has further implications for on-farm management of perennial ryegrass, as head emergence tends to both increase total herbage accumulation rate and decrease forage digestibility (Parsons 1988). In forage seed production, late heading in perennial ryegrass is reported to reduce floret fertility (Anslow 1963) and also the numbers of spikelets and florets per spike (Hill and Watkin 1975). These reports support the current results, except that in our study the later heading parent (I), which had reduced seed yield as a result of reduced FSp and SpktSp, showed a compensatory increase in FSUtil (Table 3).

Secondly, uniformity of heading within a plant, measured in this data set as spread of heading (SOH), is vital in seed production to minimize seed shattering losses of early-emerging heads and immaturity at harvest of late-maturing heads, and thereby make harvest date determinations easier. The recommended practice for managing harvesting time in commercial grass seed production is to harvest seeds at an average seed moisture content of 45% (Silberstein et al. 2010). This practice however, may be associated with a substantial reduction in seed yield of late or early heading plants. Our results show that with reduced SOH, increased FSUtil and SYPlant is also expected (Tables 4 and 5). This adds further importance to the goal of developing cultivars with synchronised heading of tillers within a population, as it not only minimises yield losses due to seed production practice but also maximises seed yield per plant. Measurement of SOH presents few logistical problems and our results revealed moderate broad sense heritability for SOH, making selection for this trait worthwhile.

Thirdly, a high number of reproductive tillers would intuitively be expected to enhance SYPlant. However, after extracting the positive effect of RTiller captured in PC1 and associated with reduced SOH, trait interactions offset much of any potential advantage from high RTiller. In particular, high RTiller

negatively affected FSUtil (Table 4; PC3 in Table 5), presumably reflecting within-plant competition factors, like the SpktSp/FSUtil and FSpkt/FSUtil trade-offs noted above. Also, total tiller number per plant pre-flowering did not translate to high RTiller (tillers from which seed-heads were harvested) or high SYPlant (data not presented). This is consistent with previous observations that increased plant tiller number in late winter-early spring does not always result in increased seed yield (Brown 1977; Hampton 1986). Differences in tiller number per plant may also reflect plant spacing and light capture strategy (Matthew et al. 1995). For instance, there was a poor correlation between seed yield of spaced and drilled plants of perennial ryegrass (Elgersma 1990a) and this was attributed to changes in the number of spikelets (Elgersma 1990a). However, spaced plants are also more likely to develop high tiller number per plant than drilled ones and differences in the production of RTiller may be another factor contributing to the poor correlation. The potential for lack of correlation between results from spaced plants and drilled trials is a salutary reminder that in any plant breeding programme trait stability across environments must be confirmed at various points in the programme.

Fourthly, selection for TSW is less likely than selection for other traits discussed above to enhance SYPlant but, where desired, could be influenced indirectly by selection for erect growth habit (Table 5), if this association proved robust across populations and environments. Measuring PGHabit is simpler than measuring TSW, and indirectly selecting plants for TSW based on erect PGHabit will enable selection of superior plants before they are harvested.

#### QTL for seed yield and related traits

Consistent with previous studies (Yamada et al. 2004; Armstead et al. 2008; Studer et al. 2008; Byrne et al. 2009; Brown et al. 2010; Paina et al. 2016), quantitative trait locus (QTL) analysis identified numerous genomic locations responsible for conditioning seed yield and seed yield-related traits. Two QTL of moderate effect were identified for SYPlant itself on LG2 and LG6, together accounting for 23% of the phenotypic variation observed in the mapping population. These co-aligned with QTL for component traits of SYPlant (SYSp, FSpkt, FSp, FSUtil, TSW, among others; Fig. 2) and these associations therefore



reflect the impact of those component traits on SYPlant. The resolution of QTL mapping does not allow for assessment of whether QTL co-linearity is due to a common genetic factor or whether there are closely-linked QTL which act independently of each other, but cumulatively. QTL of this type may well be responsible for the various independent trait associations identified in successive PCs by PCA. Similarly, QTL for SYSp on LG5 and LG7 co-locate with QTL for FSUtil, but in this case without involvement of other traits.

As with earlier studies (Yamada et al. 2004; Armstead et al. 2008; Studer et al. 2008; Byrne et al. 2009; Paina et al. 2016) multiple QTL were detected for DTH, including a QTL of large effect on LG2 (28–29%  $V_p$ ), and these were, with one exception, stable across years and environments (see below). It seems likely that these, as well as QTL for other seed yield-related traits, are the same QTL as those detected in previous reports (Supplementary Table 2) but it is difficult to confirm this definitively. In silico comparison of ryegrass QTL from different studies, based on physical positions of marker sequences in the rice genome was limited by availability of DNA sequence information for many of the QTL-linked markers. Amongst the QTL that could be compared in this way, the DTH locus on LG7 (Hd3a described by Armstead et al. 2004) and the SpLen QTL on LG2 (Brown et al. 2010) appear to correspond positionally with equivalent QTL in those studies. QTL-linked markers for DTH from both studies mapped by BLASTN to an interval of 2.7–2.9 Mb on rice chromosome 6 and for SpLen 29.7–31.4 Mb on chromosome 4.

Co-location of QTL within the current study also occurred for trait combinations that would not logically be expected to be interdependent, one example being the QTL for FSpkt, FSp, DTH, and PGHabit at 61 cM on LG4 (Table 6, Fig. 2). As per above, this may be explained either by pleiotropic effects, or by close linkage of two different genes. Another example from our data that fits this scenario is that QTL for the vegetative traits leaf appearance interval and ligule appearance interval in the same mapping population (Sartie et al. 2011) occur at or in close proximity to the LG6 SYPlant QTL (Table 6; Fig. 2). A canonical correlation analysis (Matthew et al. 1994) to explore links between the vegetative trait phenotypic data and these reproductive trait data found there was indeed a highly significant statistical association between leaf

appearance interval and SYPlant, explaining a portion of the variance for those traits (Supplementary Table 1). Further linkages between vegetative traits and heading date were also observed. Pauly et al. (2012) noted that the LG7 DTH QTL has been associated in other mapping populations with vegetative traits, including herbage yield, plant height and leaf lamina length; and proposed there may be a link between heading date and traits associated with plant growth, focused on the genes Hd3 and Hd1. Our data support this relationship, with the  $I \times S$  LG7 QTL for DTH co-aligning with QTL for leaf elongation rate and leaf elongation duration (Sartie et al. 2011). In our current study, QTL for FSUtil and SYSp were also identified in this LG7 region (Table 6; Fig. 2). Furthermore, we note that the DTH QTL on LG2 occurs at positions previously associated in population  $I \times S$  with QTL for dry matter yield (Sartie et al. 2011) and plant growth (Faville et al. 2012). The LG4 DTH QTL also aligns closely with QTL for leaf morphogenetic (Sartie et al. 2011) and field growth traits (Faville et al. 2012). As a point for future research, there is now considerable emerging understanding of the genetic control of inflorescence development and the associated molecular systems (Bommert et al. 2005; Kellogg 2007) and, with the emergence also of new genomic sequence resources for ryegrass (Byrne et al. 2015), it would be of interest to identify molecular mechanisms linked to particular QTL, especially those implicated in regulation of intra-plant competition, such as the SpktSp, FSpkt and TSW loci.

#### $G \times E$ interactions for seed yield and related traits

In their quest for understanding phenotypic variation via integrated approaches in the field environment, Pauli et al. (2016), recognised that physiological traits, plant developmental phases and growth conditions are interrelated, and this relationship requires an understanding of genotype-by-environment ( $G \times E$ ) interactions. Such interactions are widely observed in perennial ryegrass for herbage yield and other agronomic traits (Jafari et al. 2003; Conaghan et al. 2008; Fé et al. 2015; Easton et al. 2015; Kerr et al. 2012). In this study, we were able to explore the influence of  $G \times E$  on seed production after measuring a subset of traits for population  $I \times S$  at two contrasting locations. Linear mixed model analysis based on trait data from

both sites showed that GxE interaction, measured as  $\sigma_{gs}^2$ , was highly influential for SYPlant (Table 3) and this was reflected in a lack of QTL for SYPlant in common between the two environments. In contrast, smaller (albeit significant) GxE interaction was detected for both SYSp and DTH and QTL were identified at similar locations on the same linkage groups (four of five QTL positions for DTH and two of six QTL for SYSp), indicating QTL that are stable across environments. Environmentally-stable QTL and their associated markers represent particularly robust candidates for development of marker-assisted breeding to improve seed yield and manipulate heading date in perennial ryegrass.

## Conclusions

This study has provided an improved understanding of component traits associated with seed yield in perennial ryegrass, which may be targeted for selection to improve seed yield potential via conventional and molecular breeding approaches. Superficially, variation in SYPlant was strongly associated with variation in RTiller and SYSp. Deconstruction of SYSp to component traits by PCA pointed to FSUtil as being of high importance to seed yield, and showed that TSW was less important than previously believed. QTL were discovered for all investigated traits except for RTiller. QTL were identified that may be used to inform further dissection of this economically significant trait to candidate gene level. QTL-linked markers may also be utilised in MAS or as fixed factors in a genomic prediction model for improving in seed yield related traits and heading date. Nevertheless, validation of QTL effects, in other environments and genetic backgrounds, should be a prerequisite for deployment in MAS for improved seed yield in perennial ryegrass.

**Acknowledgements** This work was supported by a Bright Futures Fellowship from the New Zealand Foundation for Research, Science and Technology (later transferred to the Tertiary Education Commission), Agricom New Zealand Ltd (now part of PGG Wrightson Seeds) and AgResearch Ltd. We thank Tom Lyons (transplanting and harvesting), Mike Hickey (transplanting), Sarah Matthew (harvesting and seed processing), Robert Southward, Mark Osborne and Peter Jessop (counting TSW) for technical assistance.

## References

- Abel S, Gislum R, Boe B (2017) Path and correlation analysis of perennial ryegrass (*Lolium perenne* L.) seed yield components. *J Agron Crop Sci* 203:338–344
- Altschul SF, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Anslow R (1963) Seed formation in perennial ryegrass. 1. Anther exertion and seed set. *J Br Grassl Soc* 18:90–96
- Armstead IP, Turner LB, Farrell M, Skot L, Gomez P, Montoya T, Donnison IS, King IP, Humphreys MO (2004) Synteny between a major heading-date QTL in perennial ryegrass (*Lolium perenne* L.) and the Hd3 heading-date locus in rice. *Theor Appl Genet* 108:822–828
- Armstead IP, Turner LB, Marshall AH, Humphreys MO, King IP, Thorogood D (2008) Identifying genetic components controlling fertility in the outcrossing grass species perennial ryegrass (*Lolium perenne*) by quantitative trait loci analysis and comparative genetics. *New Phytol* 178:559–571
- Barrett B, Baird I, Woodfield D (2008) White clover seed yield: a case study in marker assisted selection. In: Yamada T, Spangenberg G (eds) Molecular breeding of forage and turf: proceedings of 5th international symposium, Sapporo. Springer, New York, pp 241–250
- Bean E (1980) Factors affecting the quality of herbage seed. In: Hebblethwaite P (ed) Seed production. Butterworths, London, pp 593–604
- Beavis W (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In: Wilkinson D (ed) Proceedings of 49th annual corn and sorghum research conference. American Seed Trade Association, Chicago, Washington, DC., pp 250–266
- Bommert P, Satoh-Nagasawa N, Jackson D, Hirano H-Y (2005) Genetics and evolution of inflorescence and flower development in grasses. *Plant Cell Physiol* 46:69–78
- Brown KR (1977) Parent seed weight, plant growth, and seeding in ‘Grasslands Tama’ Westerwolds ryegrass. *NZ J Expt Agric* 5:143–146
- Brown RN, Barker RE, Warnke SE, Cooper LD, Brilman LA, Rouf Mian MA, Jung G, Sim SC (2010) Identification of quantitative trait loci for seed traits and floral morphology in a field-grown *Lolium perenne* × *Lolium multiflorum* mapping population. *Plant Breed* 129:29–34
- Bugge G (1987) Selection for seed yield in *Lolium perenne* L. *Plant Breed* 98:149–155
- Burton GW, DeVane EH (1953) Estimating heritability in tall fescue (*Festuca arundinacea*) from replicated clonal material. *Agron J* 45:478–481
- Byrne S, Guiney E, Barth S, Donnison I, Mur LAJ, Milbourne D (2009) Identification of coincident QTL for days to heading, spike length and spikelets per spike in *Lolium perenne* L. *Euphytica* 166:61–70
- Byrne SL, Nagy I, Pfeifer M, Armstead I, Swain S, Studer B, Mayer K, Campbell JD, Czaban A, Hentrup S, Panitz F, Bendixen C, Hedegaard J, Caccamo M, Asp T (2015) A

- synteny-based draft genome sequence of the forage grass *Lolium perenne*. *Plant J* 84:816–826
- Casler MD, Barker RE, Brummer EC, Papadopolous YA, Hoffman LD (2003) Selection for orchardgrass seed yield in target versus nontarget environments. *Crop Sci* 43:532–538
- Chastain TG, Young WC, Gardback CK, Silberstein TB (2003) Seed partitioning and yield responses to trinexapac-ethyl in perennial ryegrass. In: Proc. fifth int. herbage seeds conf. International Herbage Seed Group, Gatton, pp 104–108
- Churchill D, Doerge R (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138:963–971
- Conaghan P, Casler MD, McGilloway DA, O’Kiely P, Dowley LJ (2008) Genotype × environment interactions for herbage yield of perennial ryegrass sward plots in Ireland. *Grass Forage Sci* 63:107–120
- Donald CM (1968) The breeding of crop ideotypes. *Euphytica* 17:385–403
- Easton HS, Jahufer MZZ, Flay C, Schmidt J, Rolston MP, Trehewey JAK, Ryan DL, Faville MJ (2015) Environment and ryegrass genetics-multisite trial of advanced breeding populations. *J NZ Grassl* 77:233–238
- Elgersma A (1985) Floret site utilization in grasses: definitions, breeding perspectives and methodology. *J Appl Seed Prod* 3:50–54
- Elgersma A (1990a) Spaced-plant traits related to seed yield in plots of perennial ryegrass (*Lolium perenne* L.). *Euphytica* 51:151–161
- Elgersma A (1990b) Seed yield related to crop development and to yield components in nine cultivars of perennial ryegrass (*Lolium perenne* L.). *Euphytica* 49:141–154
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:1–10
- Faville MJ, Jahufer MZZ, Hume DE, Cooper BM, Pennell CGL, Ryan DL, Easton HS (2012) Quantitative trait locus mapping of genomic regions controlling herbage yield in perennial ryegrass. *NZ J Agric Res* 55:263–281
- Fé D, Pedersen MG, Jensen CS, Jensen J (2015) Genetic and environment variation in a commercial breeding programme of perennial ryegrass. *Crop Sci* 55:631–640
- Hampton JG (1986) The effect of seed and seed lot thousand seed weight on vegetative and reproductive yields of Grasslands Moata tetraploid Italian ryegrass (*Lolium multiflorum* Lam.). *NZ J Expt Agric* 14:13–18
- Hayes BJ, Cogan NOI, Pembleton LW, Goddard ME, Wang J, Spangenberg GC, Forster JW (2013) Prospects for genomic selection in forage plant species. *Plant Breed* 132:133–143
- Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. *Crop Sci* 49:1–12
- Hill MJ, Watkin BR (1975) Seed production studies on perennial ryegrass, timothy and prairie grass. *Grass Forage Sci* 30:63–71
- International Seed Testing Association (2004) International rules for seed testing seed science and technology
- Jafari A, Connolly V, Walsh EJ (2003) Genetic analysis of yield and quality in full-sib families of perennial ryegrass (*Lolium perenne* L.) under two cutting managements. *Ir J Agric Food Res* 42:275–292
- Jin Q, Rowarth J, Scott W, Sedcole J (1996) A study of nitrogen use in a browntop (*Agrostis capillaris* L.) seed crop. *J Appl Seed Prod* 14:31–39
- Kellogg EA (2007) Floral displays: genetic control of grass inflorescences. *Curr Opin Plant Biol* 10:26–31
- Kerr GA, Chapman DF, Thom ER, Matthew C, Linden AVD, Baird DB, Johnston E, Corkran JR (2012) Evaluating perennial ryegrass cultivars: improving testing. *Proc NZ Grassl Assoc* 74:127–136
- Knott S, Neale D, Sewell M, Haley C (1997) Multiple marker mapping of quantitative trait loci. *Theor Appl Genet* 94:810–820
- Marshall AH, Wilkins PW (2003) Improved seed yield in perennial ryegrass (*Lolium perenne* L.) from two generations of phenotypic selection. *Euphytica* 133:233–241
- Matthew C, Lawoko CRO, Korte CJ, Smith D (1994) Application of canonical discriminant analysis, principal component analysis, and canonical correlation analysis as tools for evaluating differences in pasture botanical composition. *NZ J Agric Res* 37:509–520
- Matthew C, Lemaire G, Sackville HN, Hernández GA (1995) A modified self-thinning equation to describe size/density relationships for defoliated swards. *Ann Bot* 76:579–587
- Meuwissen T, Hayes B, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Paina C, Byrne SL, Studer B, Rognii OA, Asp T (2016) Using a candidate gene-based genetic linkage map to identify QTL for winter survival in perennial ryegrass. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0152004>
- Parsons AJ (1988) The effect of seasons and management on the growth of grass swards. In: Michael B, Lazenby A (eds) *The grass crop: the physiological basis of production*. Chapman and Hall, London, pp 129–177
- Pauli D, Chapman SC, Bart R, Topp CN, Lawrence-Dill CJ, Poland J, Gore MA (2016) The quest for understanding phenotypic variation via integrated approaches in the field environment. *Plant Physiol* 172:622–634
- Pauly L, Flajoulot S, Garon J, Julier B, Beguier V, Barre P (2012) Detection of favorable alleles for plant height and crown rust tolerance in three connected populations of perennial ryegrass (*Lolium perenne* L.). *Theor Appl Genet* 124:1139–1153
- Rowarth JS, Sanders KJ (1996) Relationship between seed quality teSpktSp and field emergence for *Lotus pedunculatus* (Cav.), *L. corniculatus* (L.) and *L. tenuis* (Willd.). *J Appl Seed Prod* 14:87–89
- Rowarth JS, Archie WJ, Cookson WR, Hampton JG, Sanders CJ, Silberstein TB, Young WC (1999) Link between nitrogen application, concentration of nitrogen in herbage and seed quality in perennial ryegrass (*Lolium perenne* L.). *J Appl Seed Prod* 17:1–6
- Rumball W, Foote AG (2008) Ryegrass (*Lolium* spp.) with branched inflorescences. *NZ J Agric Res* 51:265–267
- Sartie AM, Easton HS, Matthew C (2009) Plant morphology differences in two perennial ryegrass cultivars. *NZ J Agric Res* 52:391–398
- Sartie AM, Matthew C, Easton HS, Faville MJ (2011) Phenotypic and QTL analyses of herbage production-related traits in perennial ryegrass (*Lolium perenne* L.). *Euphytica* 182:295–315

- Sewell MM, Bassoni DI, Megraw RA, Wheeler NC, Neale DB (2000) Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). I. Physical wood properties. *Theor Appl Genet* 101:1273–1281
- Sewell MM, Davis MF, Tuskan GA, Wheeler NC, Elam CC, Bassoni DI, Neale DB (2002) Identification of QTLs influencing wood property in loblolly pine (*Pinus taeda* L.). II. Chemical wood properties. *Theor Appl Genet* 104:214–222
- Silberstein T, Mellbye M, Chastain T, Young W (2010) Using seed moisture as a harvest management tool. Seed Production Research Report. Oregon State University
- Sparnaaij L, Bos I (1993) Component analysis of complex characters in plant breeding. 1. Proposed method for quantifying the relative contribution of individual components to variation of the complex character. *Euphytica* 70:225–235
- Studer B, Jensen LB, Hentrup S, Brazauskas G, Kolliker R, Lubberstedt T (2008) Genetic characterisation of seed yield and fertility traits in perennial ryegrass (*Lolium perenne* L.). *Theor Appl Genet* 117:781–791
- Van Ooijen JW, Boer MP, Jansen RC, Maliepaard C (2002) MapQTL4.0, Software for calculation of QTL positions on genetic maps. International, P.R., Wageningen, The Netherlands
- White JGH (1990) Herbage seed production. In: Langer RHM (ed) Pastures: their ecology and management. Oxford University Press, Oxford, pp 370–408
- White TL, Hodge GR (1989) Predicting breeding values with applications in forest tree improvement. Forestry sciences. Kluwer Academic, Boston
- Yamada T, Jones E, Cogan I, Vecchies A, Nomura T, Hisana H, Shimamoto Y, Smith K, Hayward M, Forster J (2004) QTL analysis of morphological, developmental, and winter hardiness-associated traits in perennial ryegrass. *Crop Sci* 44:925–935