GNI-A1 mediates trade-off between grain number and grain weight in tetraploid wheat

Guy Golan[1], Idan Ayalon[1], Aviad Perry[1], Gil Zimran[1], Toluwanimi Ade-Ajayi[2], Assaf Mosquna[1], Assaf Distelfeld[2] and Zvi Peleg[1]

[1.] The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture, The Hebrew University of Jerusalem, Rehovot 7610001, Israel

[2.] School of Plant Sciences and Food Security, Tel Aviv University, Tel Aviv 6997801, Israel.

Correspondence: Zvi Peleg, The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture, The Hebrew University of Jerusalem. P.O. Box 12, Rehovot 7610001, Israel.
Tel: +972-8-9489638; Email: zvi.peleg@mail.huji.ac.il

Abstract

Grain yield is a highly polygenic trait determined by the number of grains per unit area, as well as by grain weight. In wheat, grain number and grain weight are usually negatively correlated. Yet, the genetic basis underlying trade-off between the two is mostly unknown. Here, we fine-mapped a grain weight QTL using wild emmer introgressions in a durum wheat background, and showed that grain weight is associated with the *GNI-A1* gene, a regulator of floret fertility. In-depth characterization of grain number and grain weight indicated that suppression of distal florets by the wild emmer *GNI-A1* allele increase weight of proximal grains in basal and central spikelets due to alteration in assimilate distribution. Re-sequencing of *GNI-A1* in tetraploid wheat demonstrated the rich allelic repertoire of the wild emmer gene pool, including a rare allele which was present in two gene copies and contained a non-synonymous mutation in the C-terminus of the protein. Using an F_2 population generated from a cross between wild emmer accessions Zavitan, which carries the rare allele, and TTD140, we demonstrated that this unique polymorphism is associated with grain weight, independent of grain number. Moreover, we showed, for the first time, that *GNI-A1* proteins are transcriptional activators and that selection in domesticated wheat targeted compromised activity of the protein. Our finding expand the knowledge of the genetic basis underlying trade-off between key yield components and may contribute to breeding efforts for enhanced grain yield.

Introduction

Wheat (*Triticum* sp.) is one of the major crops grown today, with production estimated at ~770 million tons per annum (http://www.fao.org/faostat). To meet the rising demand of the projected population growth by 2050, an increase of at least 60% in wheat production is required (Ray et al. 2013). Grain yield is a multifactorial trait determined by grain number (GN) and grain weight (GW), two primary yield components which are usually negatively correlated (Acreche and Slafer 2006; Bustos et al. 2013; García et al. 2013; Miralles and Slafer 1995; Sadras 2007; Slafer and Miralles 1993). GN is largely determined by the fertility of each floret within a spikelet (i.e. floret fertility). Following initiation of floret primordia, a large proportion of the florets will undergo degeneration through a genetically controlled, environment-responsive process of floret abortion (Ferrante et al. 2012; Ghiglione et al. 2008; Guo et al. 2017; Miralles et al. 1998; Ochagavía et al. 2018; Prieto et al. 2018; Sakuma et al. 2019), which defines the final number of grains per spikelet. From an evolutionary perspective, this process, which promotes high plasticity in floret/grain number, allows variable resource availability and a relatively stable phenotypical GW range (Sadras 2007; Sadras and Denison 2016). While GW is under the control of a complex genetic network expressed at various developmental stages, it is considered a very stable yield component with relatively high heritability. Final GW is largely affected by the duration and rate of linear grain growth and is a result of the interplay between potential grain weight (sink) and the actual supply of assimilates per grain during grain filling (source) (Fischer 2011).

Crop plants harbor only a small portion of the intra-species genetic diversity. The reduced genetic diversity in domesticated plants as compared to that of their progenitors (e.g. (Harlan 1992; Haudry et al. 2007) is due to the limited number of founder genotypes during domestication (Eyre-Walker et al. 1998; Mayr 1942), coupled with subsequent selection for agronomic traits, referred to as evolution under domestication (Abbo et al. 2014; Ladizinsky 1998). Experimental data have shown that GW increased during wheat evolution from its direct progenitor, wild emmer [*T. turgidum* ssp. *dicoccoides* (Körn.) Thell.] (Abbo et al. 2014; Gegas et al. 2010; Golan et al. 2015). In parallel, an increase in floret fertility (i.e., number of grains per spikelet) was evident (Sakuma et al. 2019), that together with the increase in GW, promoted grain production in modern wheat. Recently, the quantitative trait locus *Grain Number Increase 1* (*GNI1*) was identified and characterized as a homeodomain leucine zipper class I (HD-Zip I) transcription factor. Reduced function mutation (N105Y) within the conserved homeodomain of *GNI-A1*, and knockdown of

GNI1 in transgenic hexaploid wheat, indicated that it is a suppressor of floret fertility. Transcript abundance of *GNI-B1* (orthologous copy on B genome) was negligible in floral organs of tetraploid and hexaploid wheat, and was suggested to be pseudogenized in the genome of ancestral *Aegilops* species (Sakuma et al. 2019).

Potential GW, defined as the intrinsic capacity of the grain to accumulate dry matter (Bremner and Rawson 1978), is relatively low in distal grain positions. Several studies indicated that the negative relationship between average GN and GW is an outcome of the high proportion of low potential GW in high yielding cultivars (Acreche and Slafer 2006; Ferrante et al. 2015; Fischer 2008; Miralles and Slafer 1995). Alternatively, Brenner and Rawson (Bremner and Rawson 1978) showed that distal grains have similar potential GW as proximal grains. In addition, removal of florets prior to grain filling increased weight of the remaining grains (Calderini and Reynolds 2000b; Fischer and HilleRisLambers 1978), suggesting a degree of growth limitation imposed by competition for insufficient source.

The physiological mechanisms underlying the trade-off between GN and GW have been intensively studied, however, little is known about the genetic basis underlying this trade-off in wheat. Here, we fine-mapped the QTL affecting GW on chromosome 2AL (Golan et al. 2015), and examined its genetic association with the *GNI-A1* gene. Detailed characterization of the phenotype was conducted to elucidate the physiological mechanisms linking the two loci in a narrow genomic region. Moreover, we used wild and domesticated tetraploid genepools to explore the allelic variation in the *GNI-A1* gene and investigate its effect on gene function. Our findings shed new light on the interaction between yield components during wheat evolution under domestication.

Materials and methods

Plant material and experimental design

QTL mapping was conducted using a RISL (Recombinant Inbred Substitution Lines) population derived from a cross between durum cultivar Langdon (LDN) and the chromosome substitution line 'DIC-2A', which harbors the 2A chromosome of wild emmer accession FA-15-3 (ISR-A) in the homozygous background of LDN (Cantrell and Joppa 1991; Joppa 1993). For fine-mapping, three RISLs, each of which harbored the DIC-2A allele of a marker defining the *GW* QTL (RISL

#4, #63, #102), were crossed to LDN and then self-pollinated to produce F_2 progenies. F_2 and subsequently F_3 generations were screened with microsatellite markers *Xgwm558*, *Xcfa2043* and *Xhbg494* to identify homozygous recombinants within the QTL region. Eighteen F_4 backcrossed recombinant lines were used for fine-mapping of the number of grains per spikelet (GPSt) and GW.

Field experiments were carried out in an insect-proof screen-house (0.27 × 0.78 mm pore size screen) at the experimental farm of the Hebrew University of Jerusalem in Rehovot, Israel (34°47′ N, 31°54′ E; 54 m above sea level). The soil at this location is brown-red, degrading sandy loam (Rhodoxeralf) composed of 76% sand, 8% silt and 16% clay. Plants were treated with pesticides to protect against pathogens and insect pests and the plots were weeded, by hand, once a week. For QTL mapping, we applied a randomized complete block design with three replicates (single plants). For fine-mapping, we applied a split plot, complete random block design (n=5); each plot (1×1 meter) contained twenty plants per line. The number of spikelets per spike, grains per spike (GPS) and GW, were obtained from a sample of three spikes and five spikes for QTL and fine-mapping, respectively; these were used to derive the GPSt.

GW characterization in F_2 was conducted using 252 progenies derived from a cross between wild emmer accessions Zavitan and TTD140. The plants were grown in Atlit experimental station (32°41′14″N 34°56′18″E; 22m above sea level) during 2017-2018, in a brown soil (composed of 18.7% sand, 22.6% silt, 57.0% clay and 1.7% organic matter). Following anthesis, spikes were covered with glassine grain bags to avoid grain dispersal. A single spike per plant was evaluated for GN and GW. Allele determination of *GNI-A1* locus was conducted using the *Xhuj11* marker (Table S1).

Physiological characterization

Detailed characterization of the *GNI-A1* effect was conducted using LDN and a backcrossed recombinant line (63-18). Spikelet removal or distal floret removal was imposed at heading in basal (lower third of the spike) or central (middle third of the spike) spikelets (see Fig. 1) Four spikes per plot (n=5) were harvested at maturity and spikelets were manually dissected to obtain number of grains per spikelet and their corresponding GW.

To examine the effect of *GNI-A1* on grain development, grains were sampled at heading, anthesis, and 8, 14, 21, 28 and 39 days after anthesis. Four basal, central and apical spikelets were

dissected from synchronously flowering spikes (*n*=3). Grains were oven dried at 60ºC for 5 days before weighing.

Linkage and QTL analysis

Linkage analyses and map construction were performed based on the evolutionary strategy algorithm included in the MultiPoint package, as previously described (Golan et al. 2015). QTL analysis was performed with the MultiQTL package (http://www.MultiQTL.com), using the general interval mapping for Recombinant Inbred Lines (RIL populations. The hypothesis that one locus on the chromosome has an effect on a given trait (H_1) was compared with the null hypothesis (H_0) that the locus has no effect on that trait. Once the genetic model was chosen, 10,000 bootstrap samples were run to estimate the standard deviation of the main parameters: locus effect, its chromosomal position, its LOD score and the proportion of explained variation (PEV).

Haplotype analysis

A haplotype analysis was conducted based on re-sequencing data of *GNI-A1* (5'←3', 600bp) from 47 wild emmer, 28 domesticated emmer and 36 durum cultivars (Table S2). Multiple sequence alignments were performed using ClustalOmega (Sievers and Higgins 2014). Phylogenic tree construction was conducted using the *phylogeny.fr* web tool (Dereeper et al. 2008).

Copy number estimation

Genomic DNA was extracted from fresh leaf tissue of individual seedlings using the C TAB protocol. qPCR was carried out, using PerfeCTa SYBR Green FastMix (Quanta Biosciences Inc.), on the PikoReal RT-PCR system (Thermo Fisher Scientific Inc.). An efficiency value of 100±10% was confirmed for both primer sets (Table S1). DNA samples were diluted to 25ng/µL, based on a standard curve of five serial dilution points. Samples were denatured at 95˚C for 3 min, followed by 40 cycles at 95˚C for 10 sec and 60˚C for 45 sec. The $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001) was used to normalize and calibrate copy number of *GNI-A1* relative to the single copy gene *TraesCS2A02G134000*.

Transcriptional activity in yeast

Coding sequences were PCR-amplified from cDNA using primers extended with EcoRI and SalI recognition sequences. Resulting PCR products were cloned into EcoRI/SalI-digested pBD-GAL4 (Clontech, CA, USA) using standard restriction ligation. Following ligation, plasmids were transformed into *Escherichia coli* (DH5a) using the heat shock method and then sequence-verified. Site-directed mutagenesis was carried out using the QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent, CA, USA) according to manufacturer's instructions. We used a mutagenic primer (5'- GCGTGTGCGGCGGGAGCCC-GGGCTCATCCTTCTCGA-3') to mutate pBD-Hap6.

All pBD plasmids were individually transformed into the Y190 yeast strain using the LiAc/SS Carrier DNA/PEG Method (Xiao 2006). pBD-X harboring clones were isolated by growth on synthetic dextrose (SD) agar plates lacking leucine. Twenty transformed clones from each line were pooled and 3µl drops were spotted on new SD–Leu plates. Following an additional growth phase, a β-galactosidase activity assay was carried out to monitor *LacZ* expression levels. Yeast were chloroform-lysed and stained, as previously described (Park et al. 2009).

Prediction of phosphorylation sites on GNI-A1 was conducted using NetPhos 3.1 (Blom et al. 1999; Blom et al. 2004).

Statistical analysis

All statistical analyses were conducted using the JMP® ver. 14 statistical package (SAS Institute, Cary, NC, USA). Principal component analysis (PCA) was used to determine the associations between yield components. PCA was based on a correlation matrix and is presented as biplot ordinations of the backcrossed recombinant lines (PC scores). Two components were extracted using eigenvalues >1.5 to ensure meaningful implementation of the data by each factor. Correlations between yield components in backcrossed recombinant lines were estimated using the Row-wise method, using mean values for each genotype. To assess the allelic effect on GPSt and GW, mean values of experimental units were analyzed using one-way ANOVA, assuming a two-tailed distribution and equal variance. Components of descriptive statistics are graphically presented in box plots: median value (horizontal short line), quartile range (rectangle) and data range (vertical long line).

Results

GNI-A1 locus regulates grain number and grain weight

In a previous study, a major QTL controlling GW was mapped on the long arm of chromosome 2A. The wild emmer allele conferred heavier grains in the background of the durum wheat cultivar Langdon and was associated with increased grain morphometrics (Fig. 3 in Golan et al. 2015). To examine how the GW locus affects grain production, grain yield components were characterized in two field experiments (14 RISL in 2016 and 18 backcrossed recombinant lines in 2017). PCA extracted two major components, explaining collectively 77% of the phenotypic variance for the backcrossed recombinant lines. Principal component 1 (PC1; X axis) explained 50.9% of the dataset variance and was loaded positively with the number of spikes per plant (SPP), spikelets per spike (SPS), grains per spike (GPS), grains per spikelet (GPSt) grain weight per plant (GWPP) and negatively loaded with GW. PC2 (Y axis) explained 26% of the variance and was positively loaded with SPP, GWPP and GW and negatively loaded with GPS, GPSt and SPS (Fig. 1a). GW was negatively correlated with GPSt and GPS ($P<10^{-4}$) with correlation coefficients being r = 0.7 and 0.6, respectively (Table S3).

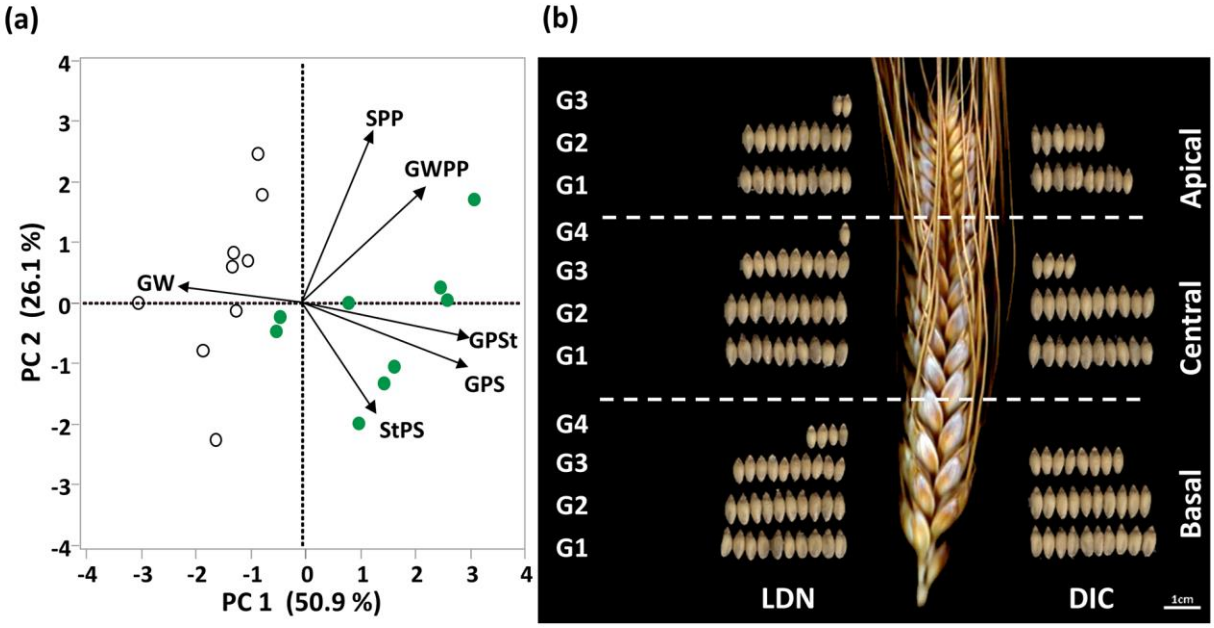

Fig. 1. The effect of grain weight QTL on grain yield components. **a)** Principal component analysis of continuous plant traits in 18 backcrossed recombinant lines. Biplot vectors are trait factor loadings for principal component (PC) 1 and PC2. Green and white dots represent genotypes carrying the domesticated (LDN) and wild (DIC) allele, respectively. Grain weight (GW), spikes per plant (SPP), grain weight per plant (GWPP), grains per spikelet (GPSt), grains per spike (GPS) and spikelets per spike (StPS). **b)** Demonstration of the locus effect on floret fertility and grain size in representative basal, central and apical spikelets.

The wild emmer (DIC) allele was associated with higher GW, while the durum (LDN) allele was associated with high GPSt and GPS (Fig. 1b, Supplementary Fig. S1). The number of spikelets per spike was not affected by the GW locus and therefore higher GPS in LDN derived directly from GPSt (r=0.96, P<0.0001). Genotypes carrying either LDN or DIC alleles had similar plant height (177 cm and 180 cm for DIC and LDN, respectively; P=0.5) and days to flowering (on average 133 days). The increase in GW by the DIC allele compensated for the lower GPS and GPSt and maintained similar grain yield between alleles in 2017. Notably, in RISL, grain GWPP was slightly higher in LDN genotypes (P=0.06) and was associated with SPP (Supplementary Fig. S1, Table S3).

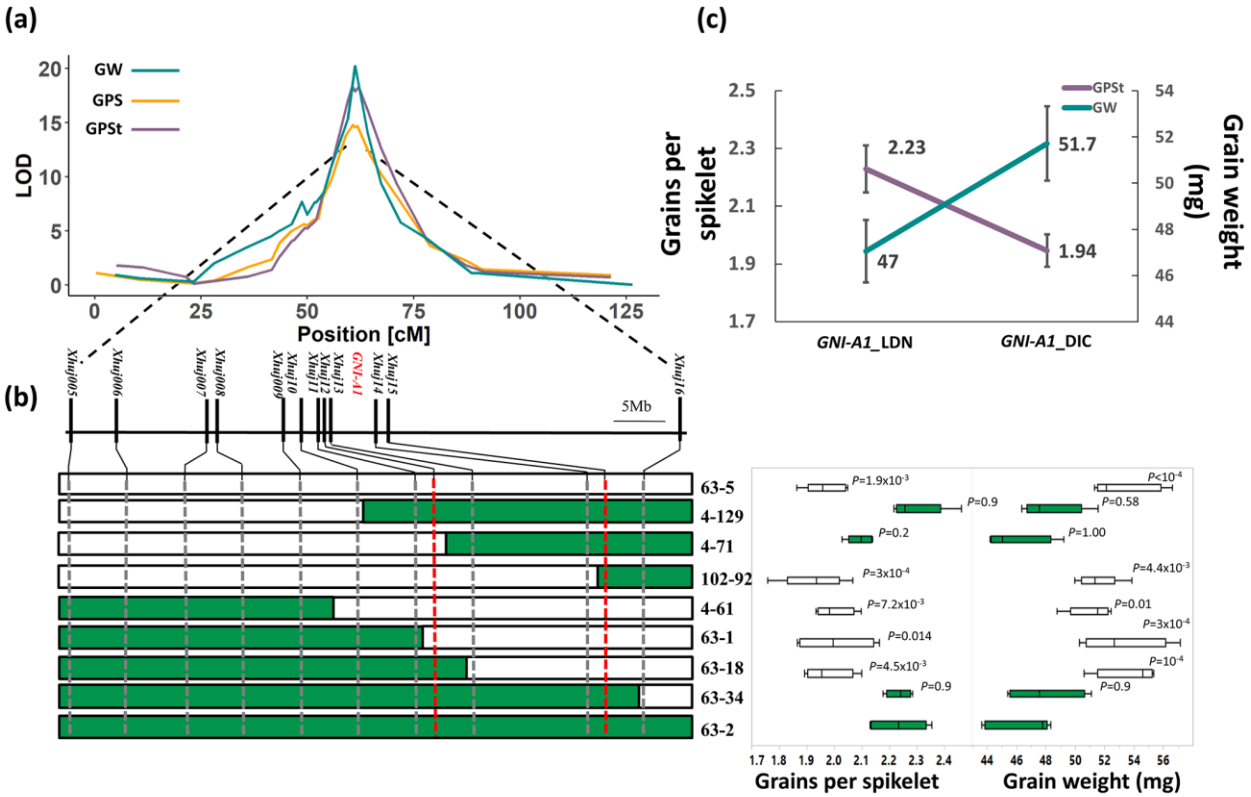

Fig. 2. Fine-mapping of grain number and weight. **a**) QTL mapping of grains per spike (GPS), grains per spikelet (GPSt) and grain weight (GW) on chromosome 2A. **b**) Graphical genotyping demonstrating fine-mapping of GW and GPSt using backcrossed recombinant lines. *P* values indicate differences from 63-2, a backcrossed recombinant line carrying domesticated allele, along the mapping interval, as determined by Dunnett's test. Red dashed lines indicate *GNI-A1* locus intervals. **c**) Significant trade-off between GPSt and GW in backcrossed recombinant lines carrying domesticated (*GNI-A1*_LDN) and wild emmer (*GNI-A1*_DIC) alleles (n=9).

Genetic analysis using a RISL population showed a stable chromosomal position of the GW QTL in two field trials (2014 and 2017), which explained 48% and 63% of the phenotypic variation, respectively (Supplementary Table S4). Subsequently, QTL mapping of GPS and GPSt showed these traits co-localize with the GW QTL and explained 54% and 61% of the phenotypic variance, respectively, (Fig. 2a, Table S4). To fine-map the QTL and dissect the genetic relationship between GPSt and GW, we used backcrossed recombinant lines and mapped the QTLs as a simple Mendelian locus. Fine-mapping delimited GW and GPSt within the 5.4 Mbp of the *GNI-A1* locus (Fig. 2b), which is associated with a significant trade-off between GPSt and GW (Fig. 2b, c). The *GNI-A1* locus contains the *GNI-A1* gene. Genotypes carrying the DIC allele harbor a functional *GNI-A1*, whereas in LDN, amino acid substitution N105Y in the conserved homeodomain, impairs gene function and increases floret fertility (Sakuma et al. 2019).

Reduction in grain number is spatially associated with enlarged grains

To study the spatial effect of the *GNI-A1* locus, GN and the weight of grains across all spikelet positions were measured in LDN and a backcrossed recombinant line (#63-18) which carries the *GNI-A1*_DIC allele in a LDN background. Basal spikelets had more grains in LDN and DIC genotypes as compared to central and apical spikelets, which carried the lowest number of grains per spikelet (Fig. S2a). Introgression of the DIC allele significantly decreased the GN in basal (2.8 *vs*. 3.3; $P<0.0001$) and central (2.0 *vs*. 2.6, $P<0.0001$) spikelets, with a non-significant effect in apical spikelets (Fig. 3a, Supplementary Fig. S2a).

The spatial effect of *GNI-A1* on GW was related to the number of grains per spikelet (Fig. 3, Supplementary Fig. S2). The DIC allele increased GW of G1 and G2 in basal and central spikelets, but had no significant effect on GW in apical spikelets and distal florets (i.e., G3) (Fig. 3b, Supplementary Fig. S2b). Characterization of grain development indicated that the effect of *GNI-A1* on GW manifested during late stages of grain filling (Supplementary Fig. S3). Moreover, the suppression of distal florets by the DIC allele decreased their proportion in the final GN (Supplementary Fig. S4), thus, contributing to the increase in average GW.

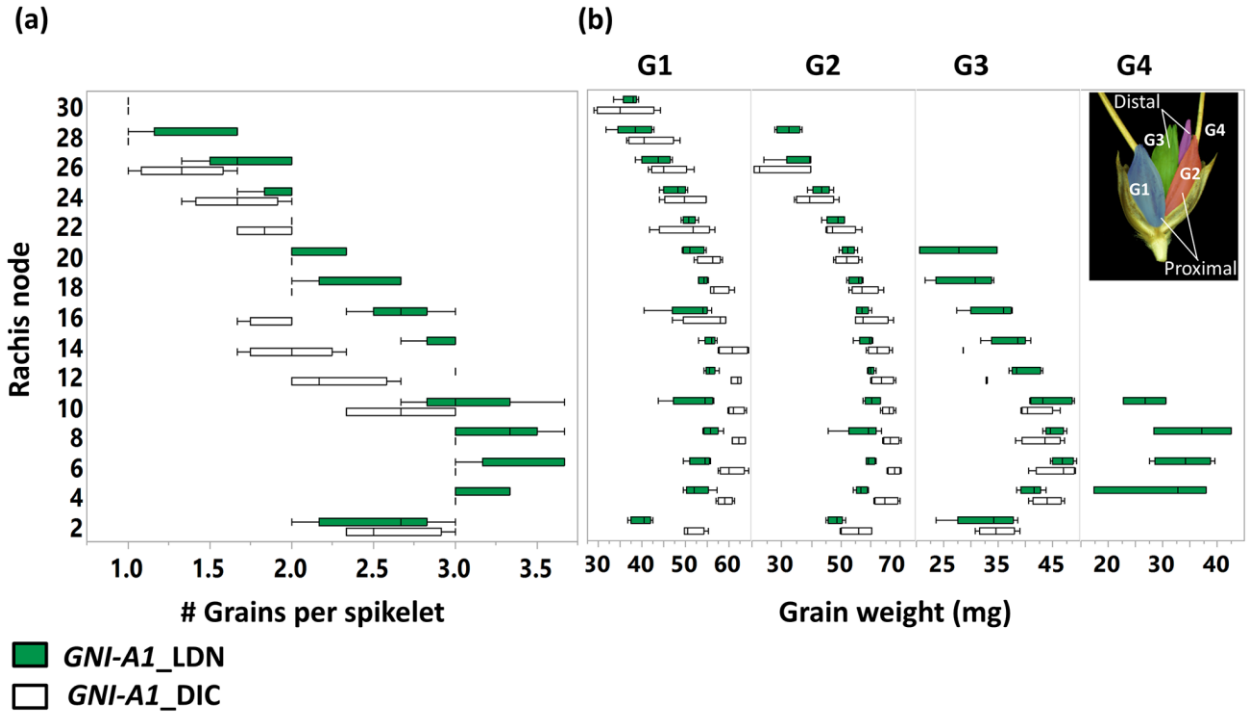

Fig. 3. Phenotypical characterization of the *GNI-A1* effect on grains per spikelet (GPSt) and grain weight (GW) in main shoots of durum cultivar Langdon (LDN) and a backcrossed recombinant line (63-18) containing the wild (*GNI-A1*_DIC) allele in the background of LDN. **a**) Number of grains per spikelet and **b**) weight of particular grain positions along the spike. Values are means of four spikes per plot (n=5). *GNI-A1_LDN* and *GNI-A1*_DIC alleles are represented by green and white box plots, respectively.

Competition among sinks is likely to reduce grain weight in LDN background

To examine if the trade-off between GN and GW in basal and central spikelets is due to competition among grains for insufficient source, we manipulated GN by removal of either basal or central spiklets, thereby altering assimilates distribution within the spike. In LDN, either basal or central spikelet removal increased weight of G1, G2 and G3 in basal, central and apical spikelets (Fig.4a, Supplementary Fig. S5a). In the backcrossed recombinant line, the effect of spikelet removal was weaker and only seldom expressed (Supplementary Fig. S5), indicating that LDN is more source-limited as compared. In order to imitate the GN phenotype in plants carrying the DIC allele, we removed only the distal florets in basal or central spikelets of LDN. Generally, removal of distal florets increased weight of G1 and G2 (Fig. 4b, Supplementary Fig.S6), indicating that a high proportion of distal grains imposes source limitation for the growth of G1 and G2. Consistent with spikelet removal, the effect of floret removal on GW was greater in LDN. Removal of distal

florets in basal spikelets increased weight of G1 (6.37 mg, *P*=0.03 *vs*. 5.00 mg, *P*=0.03) and G2 (5.25 mg, *P*=0.02 *vs*. 4.00 mg, *P*=0.4) in LDN and DIC, respectively. Removal of florets in central spikelets increased weight of G1 and G2 in LDN (5.87 mg, *P*=0.07 and 5.00 mg, *P*=0.04, respectively) but had no significant effect on GW in the backcrossed recombinant line. The effect of floret removal was not restricted to treated spikelets and extended to remote spikelets, suggesting that presence of G3 was not a physical restraint on growth of G1 and G2 (Supplementary Fig.S6). Overall, our findings suggest that lower spikelet fertility associated with the *GNI-A1*_DIC allele, eases competition among developing grains and increases proportion of high-weight grains.

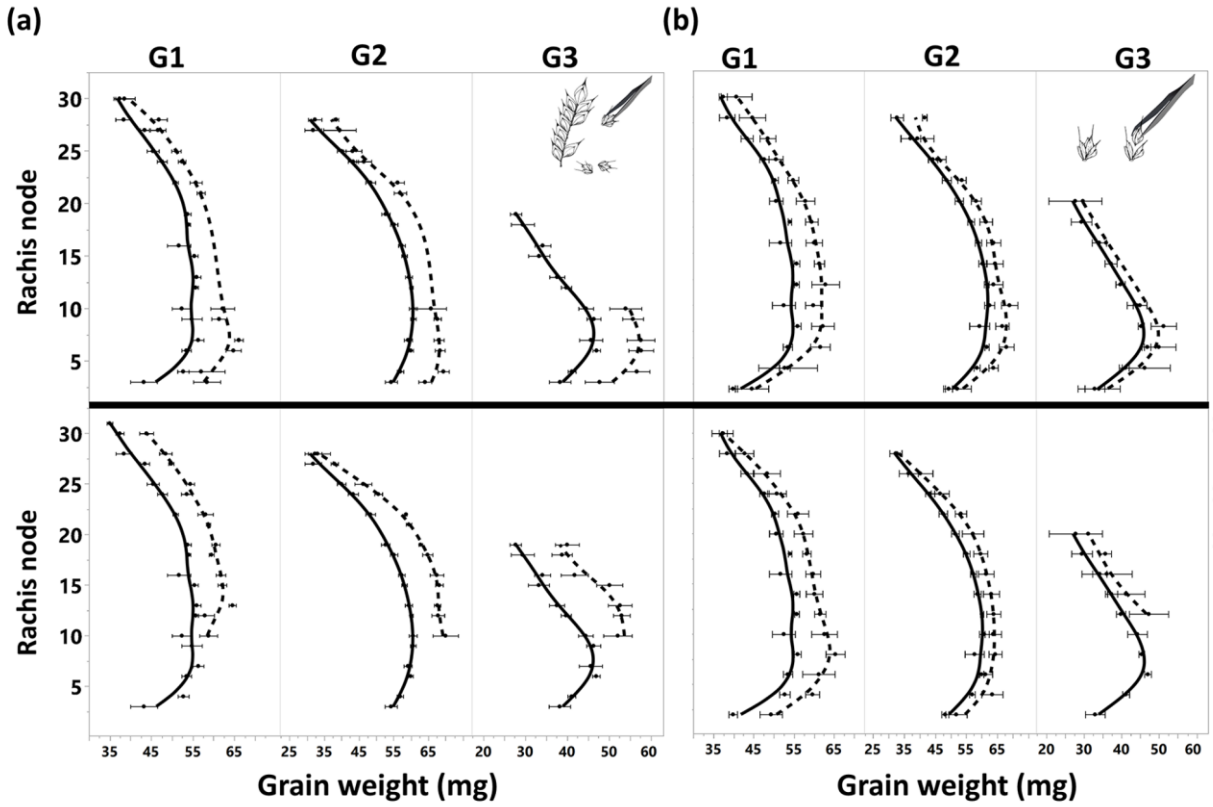

Fig. 4. Effect of source/sink modifications and spikelet position on weight of particular grain positions along the LDN spike. **a)** GW following removal of basal (lower) or central (upper) spikelets. **b)** GW following removal of distal florets in basal (lower) or central (upper) spikelets. Solid and dashed lines represent control and treated spikes, respectively. Values are mean±SE (*n*=5).

Allelic variation of *GNI-A1*

To study the origin and distribution of *GNI-A1* gene variants during wheat evolution, we re-sequenced the alleles present in 111 accessions of wild emmer, domesticated emmer and durum

wheat. We identified eight haplotypes of wild emmer and two haplotypes of domesticated emmer and durum wheat, each. Four amino acid substitutions were found among haplotypes, including the N105Y reduced-function mutation within the conserved homeodomain, that exist only in durum wheat and was shown to increase floret fertility (Sakuma et al. 2019). Three non-synonymous amino acid substitutions (i.e., N166D, S180N and G182R) were found within the C-terminal region (CTR) (Fig. 5).

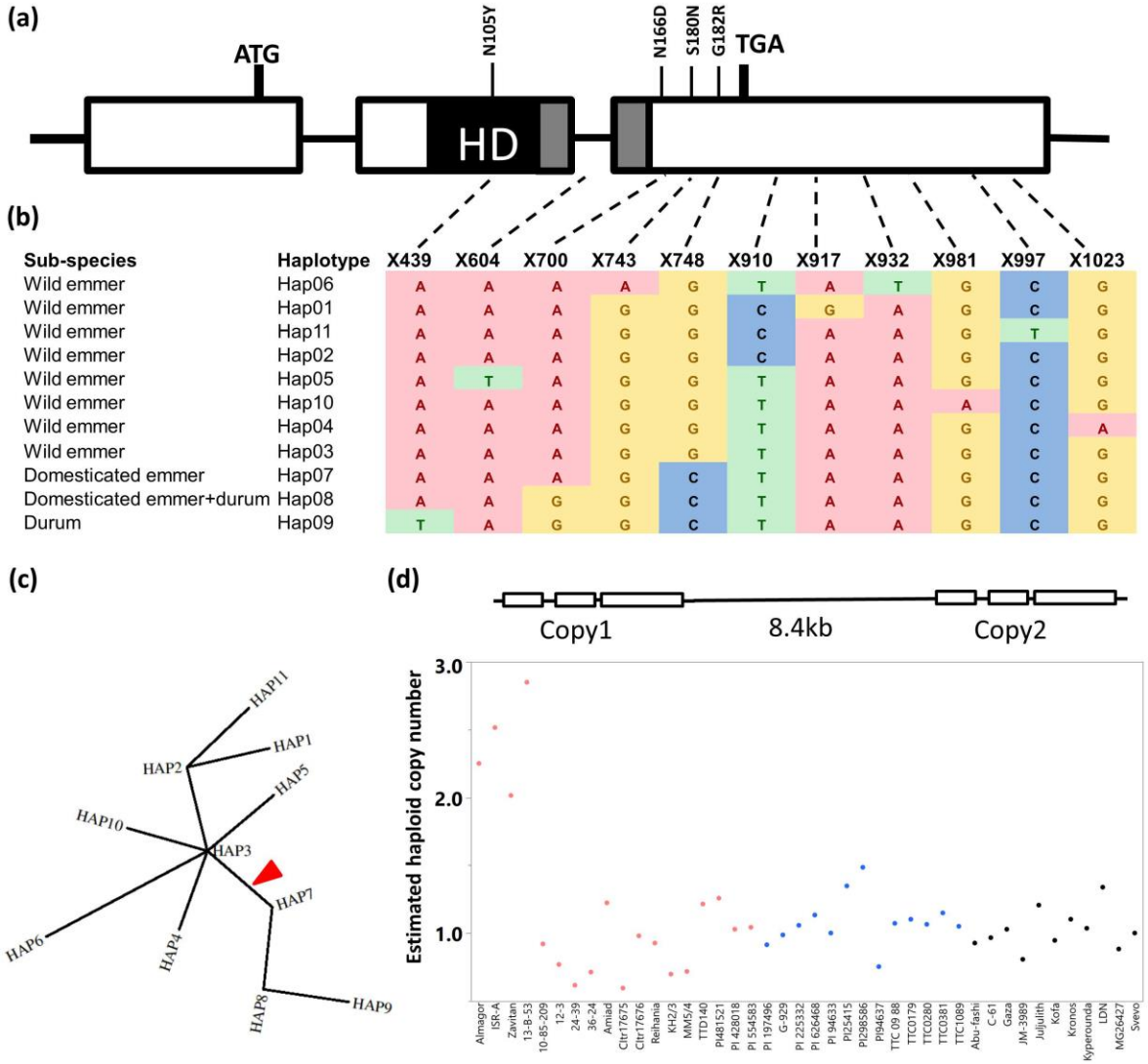

Fig. 5. Haplotype analysis of *GNI-A1* in tetraploid wheat. **a**) Gene structure of *GNI-A1*. Exons are indicated as boxes. Variation in amino acid sequence is indicated above. Single Nucleotide Polymorphism (SNP) are connected to the haplotype table by dashed lines. **b)** Eleven haplotypes classified by SNP variations were detected in tetraploid wheat. **c**) Phylogenetic analysis of the eleven haplotypes. Red triangle indicates wheat domestication on the phylogenetic tree. **d**) *GNI-A1* locus on chromosome 2AL of the wild emmer genome (accession Zavitan). Mean (n=3) qPCR estimates of *GNI-A1* haploid copy number based on *GNI-A1* quantity relative to *TraesCS2A02G134000*. Wild emmer wheat (Pink), domesticated emmer (blue) and durum wheat (black).

Haplotype analysis revealed that the S180N substitution is exceptional in wheat, confined to four wild emmer accessions from Northern Israel and absent from domesticated wheat genepool. The additional amino acid substitutions N105Y, N166D and G182R were evident only in domesticated wheat. G182R was introduced during the transition from Hap3 (wild emmer) to Hap7 (domesticated emmer) and was associated with the domestication of tetraploid wheat. N166D was introduced post-domestication (Hap8) and is present in domesticated emmer as well as durum cultivars (Fig. 5, Supplementary Table S2).

Investigation of *GNI-A1* in the wild emmer wheat genome (accession Zavitan; (Avni et al. 2017) identified two copies of *GNI-A1* on the long arm of chromosome 2A (Fig. 5d). However, in the genomes of domesticated hexaploid wheat (*cv*. Chinese Spring; https://www.wheatgenome.org) and tetraploid durum wheat (*cv*. Svevo; https://www.interomics.eu/durum-wheat-genome), we detected only a single copy of *GNI-A1*. To further characterize variation in copy number among tetraploid genepool, we estimated copy number of *GNI-A1* in 44 accessions of wild and domesticated wheat using qPCR. A single copy of *GNI-A1* was estimated in all wild and domesticated accessions tested, with the exception of four wild emmer accessions from Northern Israel, corresponding to Hap6 (accessions: Zavitan, ISR-A, 13-B-53 and Alm-1; Supplementary Table S2), which were estimated to carry two copies (Fig. 5d).

Allelic variation analysis reveals new sources for functional diversity of *GNI-A1*

In two-rowed barley, amino acid substitution (S184G) within the CTR of VRS1 (GNI-A1 barley homolog) alters phosphorylation potential of VRS1 and increases GW (Sakuma et al. 2017). Allelic variation analysis exposed a non-synonymous substitution (S180N) in a similar domain of GNI-A1, confined to wild emmer accessions (Hap6). (Figs. 5, 6a, Supplementary Table S2). To test whether the S180N amino acid substitution drives an increase in GW, we examined the association between the S180, N180 alleles and GW of 124 F_2 genotypes homozygous for *GNI-A1*. The F_2 population was developed from a cross between two wild emmer accessions, Zavitan (Hap6) and TTD140 (Hap4), both carrying a functional homeodomain (105N). Zavitan carries asparagine (N) in position 180 of GNI-A1 instead of canonical serine (S) found in TTD140 (Fig. 6a). Zavitan had more spikelets per spike and a higher number of grains per spike. GW did not differ significantly between parental lines (Fig. S7). Genetic analysis of spikes from F_2 showed

that the Zavitan allele is associated with higher GW (39 *vs*. 36 mg), regardless of grain number per spike (Fig. 6b,c). Prediction of phosphorylation sites on GNI-A1 showed that the S180N substitution resulted in hypo-phosphorylation (Supplementary Fig. S8) of the CTR.

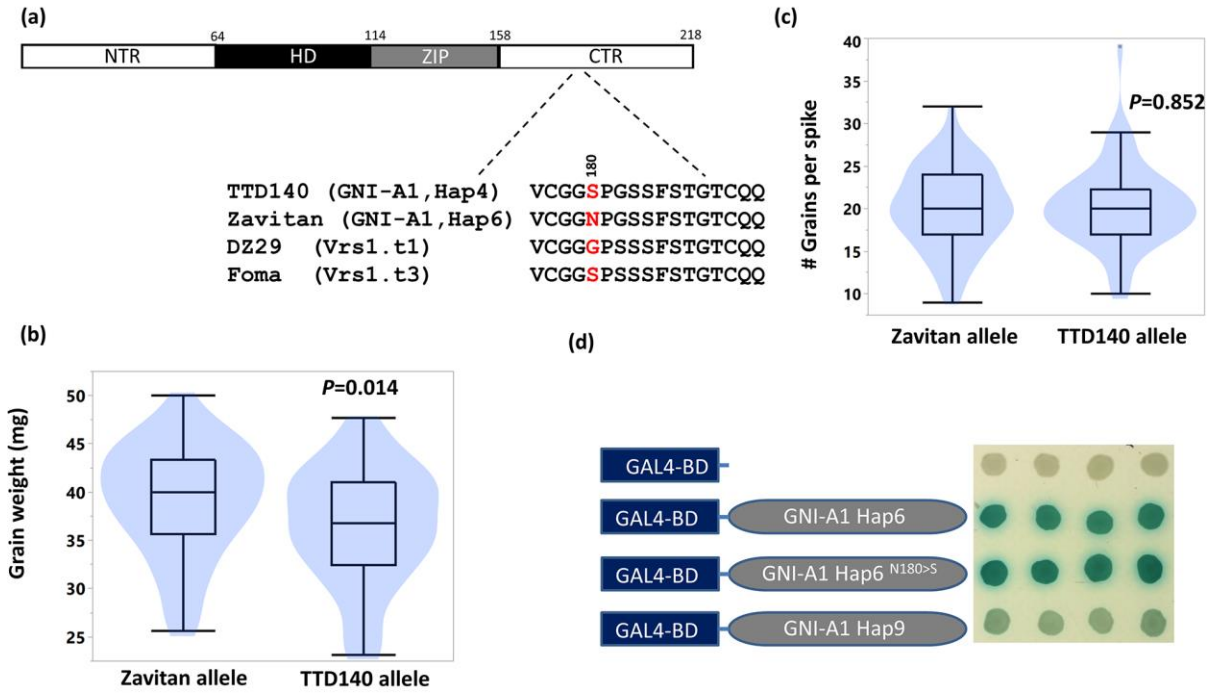

Fig. 6. Functional analysis of amino acid substitution S180N in *GNI-A1*. **a)** Wild emmer accession Zavitan and DZ29 (two-rowed barley) encode an amino acid substitution from canonical serine (S) present in wild emmer TTD140 and barley *cv*. Foma, respectively. **b**) Association between the *GNI-A1* allele and grain weight and **c**) grain number per spike, in homozygous F_2 lines derived from a cross between Zavitan and TTD140 (n=62). *P* values were determined using Student's t-test. **c**) Transcriptional activity of *GNI-A1* in yeast. The *LacZ* gene reports transcriptional activity of the Zavitan allele (Hap6), a mutated Hap6 allele, where 180N was substituted to 180S, and a domesticated durum allele (LDN, Hap9).

Reasoning that GNI-A1 might retain its activity in a heterologous system, a transcription activation assay was used to compare GNI-A1 alleles. The yeast GAL4 DNA-binding domain (GAL4BD) was fused to the open reading frame (ORF) of the wild emmer 180N allele (i.e., Zavitan, Hap6) and to a mutated version of this ORF; site-directed mutagenesis was conducted to substitute 180N to 180S. Furthermore, GAL4BD was fused to the ORF of durum cultivar LDN (Hap9) to examine the effect of mutations accumulated during domestication and further breeding. In the Y190 yeast strain, the *LacZ* reporter-gene is placed under a GAL1 upstream activating

sequence (UAS), which is recognized by BD. While the unfused BD construct (negative control) didn't induce reporter gene expression, transcriptional activity of GNI-A1 was evident in domesticated as well as in wild emmer proteins, implying that GNI-A1 is a transcriptional activator. Amino acid substitution at position 180 did not bear a substantial effect on transcriptional activity, suggesting that, at least in yeast, this substitution does not impact transcriptional activity of GNI-A1. Nevertheless, wild emmer alleles conferred higher transcriptional activity as compared to the domesticated durum allele (Fig. 6d), suggesting that during wheat evolution, selection led to decreased transcriptional activity of GNI-A1.

Discussion

Wheat grain yield is a highly polygenic trait, with low heritability, influenced by various genetic, environmental and management factors (i.e., G×E×M) and is the outcome of the interaction between grain number per m^2 and grain weight. Although the physiological mechanisms underlying trade-off between GN and GW have been extensively studied, the genetic basis underlying this trade-off is mainly unknown. Recently, evaluation of 407 winter wheat cultivars for grain and spike characteristics identified geographic as well as temporal trends showing a continuous increase in GPS due to an increase in spikelet fertility. The increase in GN associated with spikelet fertility was negatively associated with GW and the trade-off was largely controlled by attached loci on chromosome 2A (Würschum et al. 2018). In light of these report, and our findings that *GNI-A1* mediates trade-off between GN and GW, we suggest that this continuous temporal trend is controlled to some extent by *GNI-A1*. The increasing frequencies of the impaired *GNI-A1* allele over time has driven an increase in GN through the relief of floret abortion and contributed to the genetic barriers limiting an increase in GW.

GW is proposed to be the outcome of an interaction between the potential GW (sink) (i.e., when the assimilate supply is not limiting grain growth), and the actual supply of assimilate during the grain filling period (source) (Fischer 2011). Removal of distal florets in LDN significantly increased the weight of proximal grains (Fig. 4b). Moreover, removal of spikelets increased weight of G3 (Fig. 4a, Supplementary Fig. S5), indicating that in the durum wheat *cv*. LDN, distal grains are also source-limited. Lower GW in LDN was spatially related to higher fertility (Fig. 3,

Supplementary Fig. S2), suggesting that in LDN, assimilate supply is shifted towards distal grains, thereby restricting filling of proximal grains.

Irrespective of source limitation, the DIC allele increased the proportion of proximal "large grains" (Supplementary Fig. S4). Several studies suggested that the negative correlation between GN and GW derives from the large proportion of "small grains" at distal positions, and is independent of any competitive relationship among developing grains (Acreche and Slafer 2006; Ferrante et al. 2017; Ferrante et al. 2015). This proposal is supported by low correlations found between final GW and starch-synthesizing enzymes (Fahy et al. 2018), which strengthens the idea that final GW is determined by developmental processes prior to grain filling (Calderini and Reynolds 2000a; Calderini et al. 2001; Simmonds et al. 2016). Yet, in the current study, carpel weight prior to anthesis was similar, indicating that increased weight of G1 and G2 associated with the DIC allele is due to reduced competition for assimilates supply during the grain filling period.

Haplotype analysis demonstrated the loss of genetic diversity during initial wheat domestication. Subsequently, *GNI-A1* accumulated non-synonymous mutations (G182R, N166D, N105Y) during the evolution of domesticated tetraploid wheat (Fig. 5), suggesting that deliberate human selection targeted GNI-A1. The reduced transcriptional activity found in durum wheat (Fig. 6d) indicates that transcriptional activity of GNI-A1 was compromised, possibly by the introduction of N166D and G182R substitutions in the CTR. A previous study in barley has shown that the transcriptional activating domain of VRS1 is localized to the CTR, with no contribution from the NTR and HD-Zip domains (Sakuma et al. 2013). Therefore, we propose early selection of domesticated wheat varieties with reduced GNI-A1activity, a trend that was strengthened with the introduction of 105Y to durum varieties released more recently.

Allelic variation of *GNI-A1* showed that all wild emmer accessions carry the functional 105N allele, indicating that the impaired allele is associated with a fitness cost. Wild emmer has relatively heavy, arrow-shaped spikelets (dispersal unit), which fall on the ground after spike disarticulation, and penetrate the dry soil (Horovitz 1998). Thus, with such limited seed dispersal capacities, lower fecundity would reduce intra-species competition and thereby increase fitness. The observed trade-off between GN and GW may suggest that suppression of floret fertility is endorsed in wild emmer due to its positive effect on grain size. Large seeds promote seedling survival and growth (Venable 1992) due to their positive effect on the seedling vigor (Jakobsson and Eriksson 2000). Seedling vigor is a key component in plant ability to compete with

neighboring plants (Weiner 1990) and, therefore, promotion of grain size by the suppression of floret fertility is likely to aid establishment of wild emmer seedlings in their natural environment.

Hap6 represents a rare allele of *GNI-A1* located in a narrow geographic region around the Sea of Galilee in Northern Israel. Accessions from this haplotype are from the *Judaicum* subpopulation, characterized by tall culms, wide leaves, wide spikes and large grains (Avni et al. 2018; Poyarkova et al. 1991). Hap6 holds unique features such as two copies of *GNI-A1* and asparagine (N) in position 180 of the protein. Copy number variations are associated with changes in gene expression (Stranger et al. 2007) and play a key role in adaptation for a broad range of environments (Díaz et al. 2012; Würschum et al. 2015; Zhu et al. 2014). The two copies of *GNI-A1* identified in Hap6 is a phenomenon evident in a definite geographical area, characterized by wide inter-annual fluctuations in water availability (e.g., 173-859 mm $year^{-1}$; (Peleg et al. 2008)). Such stochastic environmental fluctuations may act as a driving force for evolution of local ecological adaptations associated with large grains (Metz et al. 2010). The higher GW associated with the Zavitan allele was not associated with reduction in GN (Fig. 6b, c), suggesting that Hap6 introduces functional diversity in *GNI-A1* that increases GW without reducing GN. However, the mechanism by which CNV and/or amino acid substitution promotes GW is yet to be discovered.

In two-rowed barley, substitution S184G in a parallel domain of VRS1 resulted in extreme suppression of lateral florets and promoted GW, by reducing sink competition from non-reproductive organs (Sakuma et al. 2017). Such a prominent phenotypical difference was not observed between the two wild emmer accessions carrying either 180S (TTD140) or 180N (Zavitan) (Fig. S7). Prediction of phosphorylation potential in VRS1 of two-rowed barley indicated hypophosphorylation of VRS1 due to the S184G amino acid substitution in the serine-rich motif of the CTR, which was suggested to prolong protein function throughout plant development (Sakuma et al. 2017). Here, we predicted hypophosphorylation of Hap6 proteins caused by the S180N substitution (Supplementary Fig. S8) and found that it does not compromise transcriptional activity of GNI-A1 in a heterologous yeast system. Considering that the assay reports only transcriptional changes that emanate from structural modifications of the protein, this finding suggests that the S180N substitution may affect GNI-A1 function through interference with regulatory processes. Moreover, the additional amino acid substitutions introduced during domesticated wheat evolution, reduced transcriptional activity and may be regarded as fundamental for the interaction of *GNI-A1* with the transcription machinery.

459 Overall, our findings expand the knowledge of the genetic basis underlying trade-off
460 between GN and GW. Exploration of the tetraploid gene pool suggests that wild emmer possesses
461 a unique *GNI-A1* allele that may facilitate breeding for grain weight without a significant
462 compromise on grain number.
463
464
465 **Acknowledgements**
466 We highly appreciate the excellent technical support of the Peleg lab members. Specially, we thank
467 N. Teboul for drawing in Figure 4. This research was supported by the Chief Scientist of the Israel
468 Ministry of Agriculture and Rural Development (grant #12-01-0005), and the U.S. Agency for
469 International Development Middle East Research and Cooperation (grant # M34-037).
470

References

Abbo S, Pinhasi van-Oss R, Gopher A, Saranga Y, Ofner I, Peleg Z (2014) Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. Trends in Plant Science 19:351-360

Acreche MM, Slafer GA (2006) Grain weight response to increases in number of grains in wheat in a Mediterranean area. Field Crops Research 98:52-59

Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, Hale I, Mascher M, Spannagl M, Wiebe K, Jordan KW, Golan G, Deek J, Ben-Zvi B, Ben-Zvi G, Himmelbach A, MacLachlan RP, Sharpe AG, Fritz A, Ben-David R, Budak H, Fahima T, Korol A, Faris JD, Hernandez A, Mikel MA, Levy AA, Steffenson B, Maccaferri M, Tuberosa R, Cattivelli L, Faccioli P, Ceriotti A, Kashkush K, Pourkheirandish M, Komatsuda T, Eilam T, Sela H, Sharon A, Ohad N, Chamovitz DA, Mayer KFX, Stein N, Ronen G, Peleg Z, Pozniak CJ, Akhunov ED, Distelfeld A (2017) Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science 357:93-97

Avni R, Oren L, Shabtay G, Assili S, Pozniak C, Hale I, Ben-David R, Peleg Z, Distelfeld A (2018) Genome based meta-QTL analysis of grain weight in tetraploid wheat identifies rare alleles of *GRF4 a*ssociated with larger grains. Genes 9:636

Blom N, Gammeltoft S, Brunak S (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of Molecular Biology 294:1351-1362

Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics 4:1633-1649

Bremner P, Rawson H (1978) The weights of individual grains of the wheat ear in relation to their growth potential, the supply of assimilate and interaction between grains. Functional Plant Biology 5:61-72

Bustos DV, Hasan AK, Reynolds MP, Calderini DF (2013) Combining high grain number and weight through a DH-population to improve grain yield potential of wheat in high-yielding environments. Field Crops Research 145:106-115

Calderini DF, Reynolds MP (2000b) Changes in grain weight as a consequence of de-graining treatments at pre- and post-anthesis in synthetic hexaploid lines of wheat (*Triticum durum* x *T. tauschii*). Functional Plant Biology 27:183-191

Calderini DF, Savin R, Abeledo LG, Reynolds MP, Slafer GA (2001) The importance of the period immediately preceding anthesis for grain weight determination in wheat. Euphytica 119:199-204

Cantrell RG, Joppa LR (1991) Genetic analysis of quantitative traits in wild emmer (*Triticum turgidum* L. var. *dicoccoides*). Crop Science 31:645-649

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Research 36:465-469

Díaz A, Zikhali M, Turner AS, Isaac P, Laurie DA (2012) Copy number variation affecting the *photoperiod-B1* and *vernalization-A1* genes is associated with altered flowering time in wheat (*Triticum aestivum*). PLOS ONE 7:e33234

Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS (1998) Investigation of the bottleneck leading to the domestication of maize. Proceedings of the National Academy of Sciences USA 95:4441-4446

Fahy B, Siddiqui H, David LC, Powers SJ, Borrill P, Uauy C, Smith AM (2018) Final grain weight is not limited by the activity of key starch-synthesising enzymes during grain filling in wheat. Journal of Experimental Botany 69:5461-5475

Ferrante A, Cartelle J, Savin R, Slafer GA (2017) Yield determination, interplay between major components and yield stability in a traditional and a contemporary wheat across a wide range of environments. Field Crops Research 203:114-127

Ferrante A, Savin R, Slafer GA (2012) Floret development and grain setting differences between modern durum wheats under contrasting nitrogen availability. Journal of Experimental Botany 64:169-184

Ferrante A, Savin R, Slafer GA (2015) Relationship between fruiting efficiency and grain weight in durum wheat. Field Crops Research 177:109-116

Fischer R, HilleRisLambers D (1978) Effect of environment and cultivar on source limitation to grain weight in wheat. Australian Journal of Agricultural Research 29:443-458

Fischer RA (2008) The importance of grain or kernel number in wheat: A reply to Sinclair and Jamieson. Field Crops Research 105:15-21

Fischer RA (2011) Wheat physiology: a review of recent developments. Crop and Pasture Science 62:95-114

García GA, Hasan AK, Puhl LE, Reynolds MP, Calderini DF, Miralles DJ (2013) Grain yield potential strategies in an elite wheat double-haploid population grown in contrasting environments. Crop Science 53:2577-2587

Gegas VC, Nazari A, Griffiths S, Simmonds J, Fish L, Orford S, Sayers L, Doonan JH, Snape JW (2010) A genetic framework for grain size and shape variation in wheat. Plant Cell 22:1046-1056

Ghiglione HO, Gonzalez FG, Serrago R, Maldonado SB, Chilcott C, Curá JA, Miralles DJ, Zhu T, Casal JJ (2008) Autophagy regulated by day length determines the number of fertile florets in wheat. The Plant Journal 55:1010-1024

Golan G, Oksenberg A, Peleg Z (2015) Genetic evidence for differential selection of grain and embryo weight during wheat evolution under domestication. Journal of Experimental Botany 66:5703-5711

Guo Z, Chen D, Alqudah AM, Röder MS, Ganal MW, Schnurbusch T (2017) Genome-wide association analyses of 54 traits identified multiple loci for the determination of floret fertility in wheat. New Phytologist 214:257-270

Harlan JR (1992) Crops and Man. Second Edition. American Society of Agronomy and Crop Science Society of America, Madison, WI

Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glemin S, David J (2007) Grinding up wheat: A massive loss of nucleotide diversity since domestication. Molecular Biology and Evolution 24:1506-1517

Horovitz A (1998) The soil seed bank of wild emmer. In: Zeneirei N, Kaya Z, Anikster Y, Adams WT (eds) Proceedings of the International Symposium on In-Situ Conservation of Plant Genetic Diversity. Central Research Institute of Field Crops., Ankara, Turkey

Jakobsson A, Eriksson O (2000) A comparative study of seed number, seed size, seedling size and recruitment in grassland plants. Oikos 88:494-502

Joppa LR (1993) Chromosome engineering in tetraploid wheat. Crop Science 33:908-913

Ladizinsky G (1998) Plant evolution under domestication. Kluwer Academic, Dordrecht, The Netherlands

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-DDCT} method. Methods 25:402-408

Mayr E (1942) Systematics and the origin of species. Columbia University Press, New York

564 Metz J, Liancourt P, Kigel J, Harel D, Sternberg M, Tielbörger K (2010) Plant survival in relation
565 to seed size along environmental gradients: a long-term study from semi-arid and
566 Mediterranean annual plant communities. Journal of Ecology 98:697-704
567 Miralles DJ, Katz SD, Colloca A, Slafer GA (1998) Floret development in near isogenic wheat
568 lines differing in plant height. Field Crops Research 59:21-30
569 Miralles DJ, Slafer GA (1995) Yield, biomass and yield components in dwarf, semi-dwarf and tall
570 isogenic lines of spring wheat under recommended and late sowing dates. Plant Breeding
571 114:392-396
572 Ochagavía H, Prieto P, Savin R, Slafer GA, Griffiths S (2018) Dynamics of floret initiation/death
573 determining spike fertility in wheat as affected by *Ppd* genes under field conditions. Journal
574 of Experimental Botany 69:2633-2645
575 Park S-Y, Fung P, Nishimura N, Jensen DR, Fujii H, Zhao Y, Lumba S, Santiago J, Rodrigues A,
576 Chow T-FF, Alfred SE, Bonetta D, Finkelstein R, Provart NJ, Desveaux D, Rodriguez PL,
577 McCourt P, Zhu J-K, Schroeder JI, Volkman BF, Cutler SR (2009) Abscisic acid inhibits
578 type 2C protein phosphatases via the PYR/PYL family of START proteins. Science
579 324:1068-1071
580 Peleg Z, Saranga Y, Krugman T, Abbo S, Nevo E, Fahima T (2008) Allelic diversity associated
581 with aridity gradient in wild emmer wheat populations. Plant, Cell & Environment 31:39-49
582 Poyarkova H, Gerechter-Amitai ZK, Genizi A (1991) Two variants of wild emmer (*Triticum*
583 *dicoccoides*) native to Israel: morphology and distribution. Canadian Journal of Botany
584 69:2772-2789
585 Prieto P, Ochagavía H, Savin R, Griffiths S, Slafer GA (2018) Physiological determinants of fertile
586 floret survival in wheat as affected by earliness per se genes under field conditions. European
587 Journal of Agronomy 99:206-213
588 Ray DK, Mueller ND, West PC, Foley JA (2013) Yield trends are insufficient to double global
589 crop production by 2050. PLOS One 8:e66428
590 Sadras VO (2007) Evolutionary aspects of the trade-off between seed size and number in crops.
591 Field Crops Research 100:125-138
592 Sadras VO, Denison RF (2016) Neither crop genetics nor crop management can be optimised.
593 Field Crops Research 189:75-83

Sakuma S, Golan G, Guo Z, Ogawa T, Tagiri A, Sugimoto K, Bernhardt N, Brassac J, Mascher M, Hensel G, Ohnishi S, Jinno H, Yamashita Y, Ayalon I, Peleg Z, Schnurbusch T, Komatsuda T (2019) Unleashing floret fertility in wheat through the mutation of a homeobox gene. Proceedings of the National Academy of Sciences USA, *In press*

Sakuma S, Lundqvist U, Kakei Y, Thirulogachandar V, Suzuki T, Hori K, Wu J, Tagiri A, Rutten T, Koppolu R, Shimada Y, Houston K, Thomas WTB, Waugh R, Schnurbusch T, Komatsuda T (2017) Extreme suppression of lateral floret development by a single amino acid change in the *VRS1* transcription factor. Plant Physiology 175:1720-1731

Sakuma S, Pourkheirandish M, Hensel G, Kumlehn J, Stein N, Tagiri A, Yamaji N, Ma JF, Sassa H, Koba T, Komatsuda T (2013) Divergence of expression pattern contributed to neofunctionalization of duplicated HD-Zip I transcription factor in barley. New Phytologist 197:939-948

Sievers F, Higgins DG (2014) Clustal Omega, accurate alignment of very large numbers of sequences. In: Russell DJ (ed) Multiple Sequence Alignment Methods. Humana Press, Totowa, NJ, pp 105-116

Simmonds J, Scott P, Brinton J, Mestre TC, Bush M, del Blanco A, Dubcovsky J, Uauy C (2016) A splice acceptor site mutation in TaGW2-A1 increases thousand grain weight in tetraploid and hexaploid wheat through wider and longer grains. Theoretical and Applied Genetics 129:1099-1112

Slafer GA, Miralles DJ (1993) Fruiting efficiency in three Bread Wheat (Tritkum aestivum) cultivars released at different eras. Number of grains per spike and grain weight. Journal of Agronomy and Crop Science 170:251-260

Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavaré S, Deloukas P, Hurles ME, Dermitzakis ET (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 315:848-853

Venable DL (1992) Size-number trade-offs and the variation of seed size with plant resource status. The American Naturalist 140:287-304

Weiner J (1990) Asymmetric competition in plant populations. Trends in Ecology & Evolution 5:360-364

Würschum T, Boeven PHG, Langer SM, Longin CFH, Leiser WL (2015) Multiply to conquer: Copy number variations at Ppd-B1 and Vrn-A1 facilitate global adaptation in wheat. BMC Genetics 16:96

Würschum T, Leiser WL, Langer SM, Tucker MR, Longin CFH (2018) Phenotypic and genetic analysis of spike and kernel characteristics in wheat reveals long-term genetic trends of grain yield components. Theoretical and Applied Genetics 131:2071-2084

Xiao W (2006) Yeast Protocols. Humana Press, Totowa, NJ

Zhu J, Pearce S, Burke A, See DR, Skinner DZ, Dubcovsky J, Garland-Campbell K (2014) Copy number and haplotype variation at the VRN-A1 and central FR-A2 loci are associated with frost tolerance in hexaploid wheat. Theoretical and Applied Genetics 127:1183-1197