ROYAUME DU MAROC
Université Mohammed V - RABAT -
جامعة محمد الخامس - الرباط -
Faculté des sciences
كلية العلوم

CENTRE D'ETUDES DOCTORALES - SCIENCES ET TECHNOLOGIES

N° d'ordre 3331

THESE

En vue de l'obtention du : ***DOCTORAT***

Structure de recherche : Équipe de Microbiologie et Biologie Moléculaire

Discipline : Biologie

Spécialité : Biotechnologie et Amélioration Génétique des Plantes

Présentée et soutenue le 18/09/2020 par :
Mme Hafssa KABBAJ

Genetic dissection of climate change responsive traits in a global durum wheat (*Triticum turgidum* sp. *durum*) collection exposed to a north-south heat gradient

Devant le jury :

Abdelkarim FILALI MALTOUF	**PES**, Membre Correspondant de l'Académie Hassan II des Sciences et Techniques, Rabat	**Président**
Loubna BELQADI	**PES**, Institut Agronomique et Vétérinaire Hassan II, Rabat	**Rapporteur/ Examinatrice**
Laila SBABOU	**PH**, Faculté des sciences, Université Mohammed V, Rabat	**Rapporteur/ Examinatrice**
Cherkaoui MOUDAFAR	**PES**, Faculté des sciences et techniques Université Cadi Ayyad, Marrakech	**Rapporteur/ Examinateur**
Zakaria KEHEL	**Expert**, International Center for Agricultural Research in the Dry Areas (ICARDA), Rabat	**Examinateur**
Miloudi NACHIT	**Expert**, International Center for Agricultural Research in the Dry Areas (ICARDA), Rabat	**Invité**
Ahmed AMRI	**Expert**, International Center for Agricultural Research in the Dry Areas (ICARDA), Rabat	**Invité**
Bouchra BELKADI	**PES**, Faculté des sciences, Université Mohammed V, Rabat	**Directeur de thèse**
Filippo Maria BASSI	**Expert**, International Center for Agricultural Research in the Dry Areas (ICARDA), Rabat	**Co-directeur de thèse**

Année Universitaire : 2019/2020

Dedication

*This thesis work is dedicated to my parents, **Ahmed**, and **Aisha**, who have always loved me unconditionally and whose good examples have taught me to work hard for the things that I aspire to achieve.*

*This work is also dedicated to my sweetheart sister, **Imane**, who has been a constant source of support and encouragement during the challenges of graduate school and life. I am truly thankful for having you in my life.*

Acknowledgements

The work presented in this manuscript was held at the international center of agricultural research in the dry areas (**ICARDA**) in collaboration with the laboratory of microbiology and molecular biology (**LMBM**) at the Faculty of Science of Rabat (FSR), University Mohammed V Rabat, under the supervision of professor **Bouchra BELKADI** and co-supervision of Doctor **Filippo MARIA BASSI**.

This work was realized as part of **'U-Forsk 2013'** project to deliver heat tolerant durum wheat germplasm to the Senegal river that was funded by the Swedish Research Council.

I would like to express my warm thanks to the laboratory of microbiology and molecular biology and the international center of research in dry areas to give me the opportunity to conduct my PhD.

A special thanks goes to Prof. **Abdelkarim FILALI MALTOUF** (PES, Corresponding Member of the Hassan II Academy of Sciences and Techniques), for being the president of the jury members.

I would like to thank, Prof. **Cherkaoui El MODAFAR** (PES Faculty Cadi Ayyad, Marrakech). To have accepted to report and examine my thesis.

Many thanks go to Prof. **Loubna BELQADI** (PES, Institut Agronomique et Vétérinaire Hassan II, Rabat) To have accepted to report and examine my thesis.

I also thank Prof. **Laila SBABOU** (PH, faculty of sciences, Rabat), To have accepted to report and examine my thesis.

I would like to thank Dr. **Miloudi NACHIT** (Expert, ICARDA) who has suggested me for this thesis project.

Many thanks to Dr. **Ahmed AMRI** (Expert, ICARDA) a head of gene bank at ICARDA for being part of this collaboration.

Thank you Dr. **Zakaria KEHEL** (Expert, ICARDA), for your guidance and help in biometrics.

I would like to thank my supervisor, Prof. **Bouchra BELKADI** (PES, faculty of sciences Rabat), for the patient guidance, encouragement, and advice she had provided throughout my time as her student

Many thanks go to Dr. **Filippo MARIA BASSI** (Expert, ICARDA), I have been extremely lucky to have supervisor who cared so much about my work, and who responded to my questions and queries so promptly.

I would like to thank Prof. **Rodomiro ORTIZ** and Dr. **Mulatu GUIETTA** from Swedish university of agriculture 'SLU', from ICARDA, whom without this work couldn't be accomplished.

I would also like to thank all the members of ICARDA staff in Morocco.

I would like to thank the Swedish Research Council thereby 'U-Forsk 2013' project, not only for providing the funding which allowed me to undertake this research, but also for giving me the opportunity to attend conferences and meet so many interesting people.

I must express my gratitude to the "durum PhD GANG" my team; **Meryem, Khaoula, Amadou, Priyanka, Mounira** and **Samir**, thanks for the nice moments we had together.

Finally, I would like to express my sincere gratitude to **my family**, who have experienced all the ups and downs of my research. Thanks for your patience and continued support and encouragement. I was continually amazed by having you on my side.

List of publications

- **Kabbaj H**, Sall AT, Al-Abdallat A, Geleta M, Amri A, Filali-Maltouf A, Belkadi B, Ortiz R and Bassi FM (2017). Genetic Diversity within a Global Panel of Durum Wheat (Triticum durum) Landraces and Modern Germplasm Reveals the History of Alleles Exchange. *Front. Plant Sci.* 8:1277. doi: 10.3389/fpls.2017.01277.

- Amadou T. Sall, Madiama Cisse, Habibou Gueye, **Hafssa Kabbaj**, Ibrahima Ndoye, Abdelkarim Filali-Maltouf, Bouchra Belkadi, Mohamed El-Mourid, Rodomiro Ortiz & Filippo M. Bassi. (2017). Heat tolerance of durum wheat (*Tritcum durum* Desf.) elite germplasm tested along the Senegal River. *Journal of Agricultural Science*. Vol. 10, No. 1.
- ZAÏM, Meryem[1], **KABBAJ, Hafssa[1]**, KEHEL, Zakaria, et al. Combining QTL analysis and genomic predictions for four durum wheat populations under drought conditions. Frontiers in Genetics, 2020, vol. 11, p. 316. ([1]:first authors)

- Gupta, P., **Kabbaj, H**., El Hassouni, K., Maccaferri, M., Sanchez-Garcia, M., Tuberosa, R., & Bassi, F. M. (2020). Genomic regions associated with the control of flowering time in durum wheat. Plants, 9(12), 1628.

- **Hafssa Kabbaj**, Amadou T. Sall, Madiama Cisse, Habibou Gueye, Ayed Al-Abdallat, Mulatu Geleta, Ahmed Amri, Abdelkarim Filali-Maltouf, Bouchra Belkadi, Rodomiro Ortiz and Filippo M. Bassi. Heat Tolerant Loci of Durum Wheat Defined via a North-South Gradient. (in press).

- El Haddad, N., **Kabbaj, H**., Zaïm, M., El Hassouni, K., Tidiane Sall, A., Azouz, M., ... & Bassi, F. M. (2020). Crop wild relatives in durum wheat breeding: Drift or thrift? Crop Science, 61(1), 37-54.

List of oral presentations

- **H. Kabbaj**, A. Tidiane Sall, A. Al-Abdallat, M. Geleta Dida, A. Filali-Maltouf, B. Belkadi, R. Ortiz, and F. Bassi. Genetic Dissection of Climate Change Responsive Traits in a Global Durum Wheat (*Triticum Durum* Desf) Collection Exposed to A North-south Heat Gradient. *1St International Students Symposium*. ICARDA-Rabat, Morocco 20 January 2017.

  Awarded as **''BEST ORAL PRESENTATION''**

- **H. Kabbaj**, A. Tidiane Sall, A. Al-Abdallat, M. Geleta Dida, A. Filali-Maltouf, B. Belkadi, R. Ortiz, and F. Bassi. Genome scan to identify valuable durum wheat germplasm for heat stress. *2St International Students Symposium*. ICARDA-Rabat, Morocco March 2018.

- Bassi F, EL Hassouni K, Alahmad Samir, **Kabbaj H** et al. The magic of crop wild relatives in durum wheat breeding. Pre-breeding utilizing Crop Wild Relatives. April 2019, ICARDA, Morocco.

- Filippo M Bassi, A Gatum, Sall AT, Benbelkacem A, Taghouti M, Amamou A6 Ferrahi M, Sourour A, Lagesse W, Cisse M, El Amil R, **Kabbaj H**, EL Hassouni K, Zaim M, Niane A, Sai Prasad V, Baum M. (2019). Concept idea: a product profile strategy to deliver the durum wheat varieties of tomorrow. In: Recent Trends in Wheat and Barley Research and Development. 58th All India Wheat and Barley Research Workers' Meet at Indore during August 24 - 26, 2019.

List of poster presentations

- **H. Kabbaj**, A. Al-Abdallat, A. Filali-Maltouf, B. Belkadi, M. Nachit, R. Ortiz, J. Poland, and F. M. Bassi. Multi-populations genetic map as a tool for the application of genomic selection in durum wheat. *International Conference "FROM SEED TO PASTA & BEYOND" Bologna*, Italy 31 May - 2 June 2015.

- **H. Kabbaj**, M. Zaim, Z. Kehel, G. Gorjanc, A. Al-Abdallat, A. Filali-Maltouf, B. Belkadi, M. Nachit, R. Ortiz, J. Poland, and F. M. Bassi. Genome scan and genomic selection to identify valuable durum wheat germplasm for abiotic stress. *International conference BORLAUG GLOBAL RUST INITIATIVE (BGRI) –* Morocco, Marrakech, April, 2018.

- **H. Kabbaj**, M. Zaim, Z. Kehel, G. Gorjanc , A. Al-Abdallat, A. Filali-Maltouf, B. Belkadi, M. Nachit, R. Ortiz, J. Poland, and F. M. Bassi. Genome scan and

genomic selection to identify valuable durum wheat germplasm for abiotic stress II. *International conference "FROM SEED TO PASTA III" Bologna*, Italy, September 2018.

- M. Zaïm, **H. Kabbaj**, A. Al-abdallat, G. Gorjanc, J. Poland, M. Nachit, A. Amri, K. F. Maltouf, F. M. Bassi. Genomic regions influencing yield stability in durum. *International conference "FROM SEED TO PASTA III" Bologna*, Italy, September 2018.

- Amadou Tidiane SALL, Madiama CISSE, Habibou GUEYE, **Hafssa KABBAJ**, Ibrahima NDOYE, Abdelkarim FILALI-MALTOUF, Bouchra BELKADI, Mickel BAUM Rodomiro ORTIZ, Filippo BASSI. Genomic scan in durum wheat reveals regions controlling adaptation to the heat-prone conditions of the Senegal river. *International conference "FROM SEED TO PASTA III" Bologna*, Italy, September 2018.

- Bassi F, EL Hassouni K, Gupta P, **Kabbaj H**, Zaim M, Sall T.A, Belkadi B, AL-Abdallat A, Amri A, Ortiz R, Baum M. How to adapt durum wheat when the environment tries everything to kill it. Borlaug Global Rust Initiative conference. April 2018, Marrakech, Morocco.

- Gupta P, **Kabbaj H**, EL Hassouni K, Frascaroli E, Petrozza A, Summerer S, Maccaferi M, Sanchez-Garcia M, Tuberosa R, Bassi F. Paving the path of durum: Identification of flowering genes and heterotic behavior for root vigor. Borlaug Global Rust Initiative conference. April 2018, Marrakech, Morocco.

- Bassi FM, El Hassouni K, Gupta P, **Kabbaj H**, Sall TA, Zaim M, Al-Abdallat A, Belkadi B, Filali-Maltouf A, Alary V, Amri A, Ortiz, Baum M. How to adapt durum wheat to climate change. The 1st International wheat congress (IWC). July 2019, Saskatoon, Canada.

Abstract

The countries of the Mediterranean basin cultivate durum wheat extensively, with the North shores of Africa growing more than 3 million hectares of it. Climate change is projected to have a significant impact on temperature and precipitation profiles in the Mediterranean basin, which in turn would affect the production of durum. A collection of 380 durum accessions was gathered and assessed for genetic diversity to reveal 10 sub-populations. The collection was then field tested in parallel in Morocco, Mauritania and Senegal. These sites are located 14 parallels away from each other, but with only 2 meridians of difference. this North-South gradient allows simulating the future climatic conditions that should hit North-West Africa by 2060. Combining phenotyping at these sites and genotyping allowed the identification of eighteen loci associated with global warming responsive genes, especially on chromosomes 1A, 3B and 7B. The identification of these critical heat tolerance loci unlocks the potential of deploying molecular techniques to quickly pyramid them into climate change-ready varieties. A pilot study was conducted to refine the mode of potential deployment of GS in durum wheat breeding on Four RILs populations. That were sown at three locations in Morocco and Lebanon for collection of yield and yield components, and where genotyped by sequencing. Our results shows that the prediction accuracy depend on many factors, interestingly, feeding the model with information on markers associated with significant QTLs increased the overall accuracy.

Key words: durum wheat, genetic diversity, heat gradient, QTLs, genomic prediction

Résumé

Les pays du bassin méditerranéen cultivent abondamment le blé dur avec le nord de l’Afrique en couvre plus de 3 millions d’hectares. Les changements climatiques devraient avoir un impact significatif sur les profils de température et de précipitations dans le bassin méditerranéen, ce qui aurait une incidence sur la production de blé dur. Une collection de 380 accessions de blé dur a été évaluée pour la diversité génétique. 10 sous-populations ont pu être identifiées. Ensuite, ils ont été testés en parallèle au Maroc, en Mauritanie et au Sénégal. Ces sites sont situés à 14 parallèles, mais avec seulement 2 méridiens de différence. Ce gradient Nord-Sud permet donc de simuler les futures conditions climatiques qui devraient toucher le Nord-Ouest africain d'ici 2060. La combinaison du phénotypage et du génotypage a permis l'identification de dix-huit loci associés à des gènes sensibles au réchauffement climatique, en particulier sur les chromosomes 1A, 3B et 7B. L'identification de ces critiques loci de tolérance à la chaleur permet le potentiel du déploiement de techniques moléculaires pour les pyramider rapidement en variétés prêtes au changement climatique. Une étude pilote a été menée pour affiner le mode de déploiement potentiel de la sélection génomique dans la sélection du blé dur, ceci sur quatre populations (RIL). Qui ont été semis au Maroc et au Liban pour la collecte du rendement et des composants de rendement. Et ont été génotypés par séquençage. Nos résultats confirment que la précision de la prédiction dépend plusieurs critères, ainsi alimenter le modèle avec des informations sur les marqueurs associés à des QTL significatifs augmentait la précision globale.

Mots-clés : Blé dur, diversité génétique, gradient thermique, QTL, prédiction génomique.

Résumé (détaillé)

Les pays du bassin méditerranéen cultivent abondamment le blé dur, avec la rive nord de l'Afrique couvrent plus de 3 millions d'hectares. Les changements climatiques devraient avoir un impact significatif sur les profils de température et de précipitations dans le bassin méditerranéen, ce qui aurait une incidence sur la production de blé dur. Une collection mondiale de 370 accessions de blé dur a été rassemblée et évaluée pour la diversité génétique. Au total, 10 sous-populations ont pu être identifiées, six constituées de cultivars modernes et quatre de variétés locales de différentes origines géographiques. Bien que la comparaison génomique entre les groupes a indiqué que le Moyen-Orient et l'Éthiopie avaient le niveau de diversité allélique le plus faible, tandis que les programmes de sélection et les variétés locales collectées en dehors de ces régions étaient les plus riches en allèles rares. De plus, l'analyse phylogénétique des variétés locales a indiqué que l'Éthiopie apparaît comme un deuxième centre d'origine du blé dur, plutôt que comme un deuxième site de domestication comme on le croyait précédemment. Ensuite, la même collection a été testée au champ au Maroc, en Mauritanie et au Sénégal. La plupart des conditions agro-environnementales, y compris l'irrigation et les types de sol, sont similaires entre les sites, Ce gradient Nord-Sud de 10°C, permet donc de simuler les futures conditions climatiques qui devraient toucher le Nord-Ouest africain d'ici 2060. La combinaison du phénotypage et du génotypage sur ces sites a permis l'identification de sept *loci* associés à des gènes sensibles au réchauffement climatique, en particulier sur les chromosomes 1A, 2B et 7B, qui ont été considérés comme stables tout au long du gradient thermique. L'identification de ces *loci* clés de tolérance à la chaleur permettrai le potentiel du développement des variétés prêtes à faire face au changement climatique. Le déploiement de la sélection génomique dans la sélection du blé dur a été testé sur quatre populations de lignées consanguines recombinantes. Au total, 576 individus ont été semés dans trois sites, deux au Maroc et un au Liban, pour la collecte des composants de rendement, puis ont été génotypés par séquençage. 3202 marqueurs polymorphes ont été utilisés pour tester des modèles statistiques, incorporant une matrice de relations, l'interaction génotype par environnement et l'interaction marqueur par environnement. L'utilisation de populations d'entraînement (TP) en relation 'sibs propre' avec la population de validation (VP) s'est avérée être la seule stratégie efficace, avec des précisions atteignant 0,35-0,47 pour le rendement en grains (GY). La réduction du nombre de marqueurs à 10% de l'ensemble et de la taille de la population d'entrainement à 20% n'ont pas engendré des changements significatifs de précision. L'utilisation des QTL comme effet fixe a montré un gain de précision significatif pour les quatre populations. Nos résultats confirment que la précision de la prédiction génomique dépend fortement de la parenté entre TP et VP, mais pas du nombre de marqueurs et de la taille de TP utilisés. De plus, le renforcement du modèle avec des informations sur les marqueurs associés aux QTLs a augmenté la précision.

Mots-clés : Blé dur, diversité génétique, changement climatique, gradient thermique, QTL, prédiction génomique.

Contents

Dedication ... i
Acknowledgements ... ii
List of publications ... iv
List of oral presentations ... v
List of poster presentations ... v
Abstract ... vii
Résumé ... viii
Résumé (détaillé) ... ix
Abbreviations ... xiii
List of figures ... xiv
List of tables ... xvii
General introduction ... 1
Aim of the study ... 4
Chapters description ... 5
Chapter I: literature review ... 6
I-1. History and origin of durum wheat ... 6
I-2. Durum wheat life cycle ... 7
I-3. Cultivation requirements of Durum wheat ... 9
I-4. Constraints of durum wheat production ... 10
I-5. Breeding methods and technologies ... 10
I-5-1. Conventional breeding and plant improvement ... 11
I-5-2. Experimental designs ... 11
I-5-3. Selection methods in breeding programs ... 14
✓ **Double haploid production ... 16**
✓ **Bulk breeding ... 17**
✓ **Backcross breeding: ... 19**
I-5-4. Genome wide association mapping (GWAS) ... 20
I-5-5. Quantitative trait locus (QTL) as a tool to detect genetic variation ... 22
I-5-6. Marker assisted selection ... 23
I-5-7. Genomic selection ... 24
Chapter II: Genetic Diversity within a Global Panel of Durum Wheat (*Triticum durum*) Landraces and Modern Germplasm Reveals the History of Alleles Exchange ... 26

Abstract 26
II-1. Introduction 27
II-2. Materials and Methods 29
II-2-1. Plant material 29
II-2-2. DNA extraction and genotyping 30
II-2-3. Data analysis 30
II-3. Results 31
II-3-1. Genotyping of a global panel of durum elites, cultivars and landraces 31
II-3-2. Population stratification 32
II-3-3. Admixture analysis by kinship 35
II-3-4. Genetic diversity among 10 sub-populations 38
II-4. Discussion 40
II-4-1. Success level of the clustering procedure 40
II-4-2. Clusters of genetic similarity 43
II-4-3. Comparison to other population stratification studies of durum wheat 49
II-4-4. Genetic diversity and the future of durum wheat breeding 49
Chapter III: Heat Tolerant Loci of Durum Wheat Defined via a North-South Gradient 51
Abstract 51
III-1. Introduction 52
III-2. Materials and methods 53
III-2-1. Plant material 53
III-2-2. Field experiment and phenotyping 54
III-2-3. Statistical analysis 54
III-3. Results 56
III-3-1. Yield components performance under north-south heat gradient 56
III-3-2. Heat susceptibility index (HSI) 61
III-3-3. Marker-trait associations (MTAs) for yield component and heat Susceptibility Index 62
III-3-4. Markers validation by KASP 64
III-4. Discussion 66
III-4-1. Performances of the panel under heat-stressed conditions 66
III-4-2. Genomic regions controlling heat tolerance in durum wheat 67
III-5. Conclusions 69
Chapter IV: Combining QTL Analysis and Genomic Predictions for Four Durum Wheat Populations Under Drought Conditions 70

Abstract **70**

IV-1. Introduction **71**

IV-2. Materials and methods **73**

IV-2-1. Plant material **73**

IV-2-2. DNA extraction and genotyping **74**

IV-2-3. Statistical analysis **74**

IV-3. Results **76**

IV-3-1. Genomic prediction: identification of the best fitting model (i, ii, iii, iv) **76**

IV-3-2. Genomic prediction: effect of reducing TP and marker size (v) **77**

IV-3-3. Genomic prediction: importance of relatedness between TP and VP (vi) **78**

IV-3-4. Genomic prediction: effect of QTL analysis on model accuracy (vi, viii) **79**

IV-4. Discussion **81**

IV-4-1. Selection of the best fitting statistical models for genomic predictions (i, ii, iii, iv) **82**

IV-4-2. Size and relatedness of the training population (v, vi) **82**

IV-4-3. Does markers number affect the predictions? (v) **83**

IV-4-4. Is there an advantage to conduct QTL analysis before genomic predictions? (vii, viii) **84**

IV-5. Conclusion **85**

General discussion 86

Conclusion and perspectives 92

References 94

Annexes 106

Abbreviations

AMOVA	Analysis of Molecular Variance	LOD	Logarithm of ODds
ANOVA	Analysis of Variance	MARS	Marker-Assisted Recurrent Selection
BIC	Bayesian Information Criterion	MAS	Marker Assisted Selection
BLUE	Best Linear Unbiased Estimator	MCH	Marchouch
BP	Before Present	MLM	Mixed Linear Models
BP	Breeding Population	MTA	Marker-Trait-Association
DAPC	Discriminant Analysis of Principal Components	NTSYS	Numerical Taxonomy and Multivariate Analysis System
DH	Double Haploid	PCA	Principal Component Analysis
DL	Day Length	PCoA	Principal Coordinates Analysis
DNA	DeoxyriboNucleic Acid	PIC	Polymorphism Information Content
DT	Drought Stress	KASP	Kompetitive Allele Specific PCR
DTH	Days To Heading	QTL	Quantitative Trait Loci
FIGS	Focus Identification of Germplasm Sources	RILs	Recombinant Inbreed Lines
GBS	Genotyping By Sequencing	SA	Sidi el Aydi
GDD	Growing Degree Days	SNP	Single Nucleotide Polymorphism
GEBV	Genomic Estimated Breeding Value	SSD	Single Seed Descent
$Gr.m^{-2}$	Number of Grain per meter square	SSR	Simple Sequence Repeat
Gr.spk	Number of Grain per spike	TBV	True Breeding Value
GS	Genomic Selection	TER	Terbol
GWAS	Genome Wide Association Mapping	TES	Tessaout
GXE	Genotype by Environment Interaction	TGW	Thousand grain weight
GY	Grain Yield	TKW	1,000 Kernel Weight
HSI	Heat Susceptibility Index	TP	Training Population
HT	Heat Tolerance	UPGMA	Unweighted Pair Group Method of Association
JSH	Jemaat SHaim	VP	Validation Population
LD	Linkage Disequilibrium	YP	Yield Potential

List of figures

Figure I- 1: The evolutionary and genome relationships between cultivated bread and durum wheats and related wild diploid grasses, showing examples of spikes and grain (p. r. shewry, 2009). 7

Figure I- 2: Life cycle of wheat. line drawing by tim f. knight. adapted from e. j. m kirby and margaret appleyard, cereal development guide, 2d ed. (warwickshire, england: national agricultural centre, 1987) 9

Figure I- 3: Pedigree selection method 15

Figure I- 4: Single seed descent (SSD) selection method 16

Figure I- 5: Double haploid production 17

Figure I- 6: Bulk breeding selection method 18

Figure I- 7: Backcross selection method 19

Figure I- 8: Genome wide association mapping principle. stage i: phenotyping, stage ii: genotyping and stage iii: genome-wide association study including statistical models, multiple-testing analyses, and software/packages for qtl and gene identification. 21

Figure I- 9: Schemes of genomic selection (gs) (left) and traditional mas for the selection of quantitative traits (right). (nakaya & isobe, 2012) 25

Figure II- 1: Country of origin of the durum wheat core collection lines (landraces in black, elite lines in grey). 30

Figure II- 2: Statistical determination of the optimum number of clusters by discriminant analysis of principal components (dapc). 33

Figure II- 3: Cluster analysis using discriminant analysis of principal components (dapc). (a) graphical representation of principal component 1 (ipc1) and 2 (ipc2) distances for 10 sub-populations within the whole panel. (b) graphical representation of ipc1 and ipc 35

Figure II- 4: Admixture kniship matrix for landraces and elite lines of 370 durum wheat individual....36

Figure II-5: Diversity in admixture among landraces by ad hoc structure analysis. (a) admixture analysis by kinship color coded following the colors of figure 3 (b) phylogenetic tree of evolutionary distances based on kinship values color coded following the colors set in figure 3....36

Figure II- 6: Geographical distribution of the coordinate of collection of the durum wheat landraces. color codes are provided following the colors set in figure 3 dashed circles indicate the two centers of origin or diversity. 38

Figure II- 7: Neighbor-joining tree based on snps data using rogers' genetic distance for 370 lines of durum wheat. genotypes are coded based on the output of dapc clustering. and clusters are defined using the vertical line in black as reference. 42
Figure II- 8: Spike morphology of the two centers of origin's landraces. (a) middle east landraces, (b) t.abyssinicum type. 48
Figure III- 1: Boxplot showing the variation in grain yield (gy), 1,000 kernel weight (tkw), grain per spike (gr.spk) and number of grains per meter square (gr.m-2) for each location (fanaye, kaedi, melk zhar and tassaout) during the two crop seasons 2014-15 and 201...57
Figure III- 2 : Clustering of seven environments s by principal component analysis (pca) using the weather and phenological data presented in table III-2. black circles indicate clusters of similar environments...60
Figure III-3 : Agro-morphological performances of the elite lines per clusters for gy, tkw and gr.spk. 60
Figure III- 4: The combined heat susceptibility index (hsi) for gy, tkw, gr.m-2 and gr.spk showing the number of heat tolerant and heat susceptible lines of each trait in the left, and the total percentage of excellent, good, medium and bad lines for all traits. 61
Figure III- 5: Circoss representation of markers-trait association for the traits GY, TKW, Gr.m-2 and Gr.spk and their HSI and the identified QTLs under heat stress……………………………………………………………………………………63
Figure III- 6: Kompetitive Allele Specific PCR (KASP) markers validation on an independent set of 94 elite lines of ICARDA tested under severe heat for grain yield and biomass. Correlation was measured between the BLUE for grain yield recorded along the Senegal River and the haplotype score. Accuracy, sensitivity, and specificity where determined …………………………………………………………..64

Figure IV- 1: Diagram of genomic selection (GS) processes starting from the training population and selection candidates continuing through to genomic estimated breeding value (GEBV)–based selection. (Heffiner et al.,2009) 72
Figure IV- 2 : Prediction accuracy for grain yield (GY) in YG, IG and DRO populations using four different statistical models. G+E: genotype + environment effect, GxE: genotype by environment interaction, GxE+M: genotype by environment interaction +markers effect; GxE+MxE: genotype by environment interaction + markers by environment interaction. The horizontal line represents the average, the square

indicates the 2nd and 3rd quartiles, the whiskers represent the 1st and 4th quartiles, the cross the median, and the dots are outliers. The letters indicated classes determined via LSD. 77
Figure IV- 3: Prediction accuracy for grain yield (A) and 1,000 kernel weight (B) using different randomly selected sub-sets of markers in decreasing order: 320 (10%), 640 (20%), 1,281 (40%), 921 (60%), 2,562 (80%), and 3,202 (100%) tested on DRO population using 20%, 50% and 75% of the whole population as training set (TP) to predict the rest of the population (VP). The whiskers represent the standard errors. 78
Figure IV- 4: Prediction accuracy for grain yield (A) and 1,000 kernel weight (B) using all markers, only markers linked to QTLs, and all markers except those identified as linked to QTLs. Whiskers represents the experiment LSD……………………………..…80
Figure IV- 5: Comparison of the prediction accuracies of grain yield (GY) for the four population YG, DRO, IG, and SW, using a model with all markers considered as random effect against a models that fixed markers underlying QTLs. The horizontal line represents the average, the square indicates the 2nd and 3rd quartiles, the whiskers represent the 1st and 4th quartiles, the cross the median, and the dots are outliers. The letters indicated classes determined via LSD…………………….……..81

List of tables

Table II- 1: Number and distribution across the 14 chromosomes (chr.) of durum wheat of polymorphic snps markers on the axiom 35k breeder's array and the 2-points lod for the subset of the 500 snps used for clustering. 32
Table II- 2: Analysis of molecular variance (amova) for stratification of a global durum wheat panel into 10 sub-populations. 33
Table II- 3: Reliability of the entries assignment based on comparison between three genetic clustering methods: dapc, structure, and neighbor-joining (details available in annex 1)..34
Table II- 4: Genetic diversity per sub-population. 39
Table III- 1: The analysis of variance of yield component trait (anova) 57
Table III- 2: Temperature data at flowering time (maximum, minimum and average), growing degree days (gdd), day length (dl) at sowing and flowering times and days to heading (dth) per location showing the difference between the two sides (morocco and senegal river)...59
Table III- 3: Comparison of grain yield (gy), 1,000 kernel weight (tkw) and number of grains per spike (gr.spk) between clusters defined in figure2 61
Table III- 4 : Quantitative trait-locus (qtl) for yield component traits across clusters and their heat susceptibility index 65
Table IV- 1: Board sense heritability (h2) for grain yield (gy) and 1,000 kernel weight (tkw) for the four population (dro, ig, yg, sw) across environments............................76

General introduction

Food security has become a major challenge given the projected need to increase world food supply by about 70% by 2050 (Anon, 2009). Considering the limitations on expanding crop-growing areas, a significant increase in crop productivity will be required to achieve this target (Parry et al., 2011; Reynolds et al., 2011). Increasing temperature and incidence of drought associated with global warming are posing serious threats to food security (Lobell et al., 2013).

Climate change is a threat to agriculture and food security and there is an urgent need to identify priorities for future research. It is also likely to lead to an increase in temperature. Climate models show a high probability (>90%) that by the end of this century, growing season temperatures will exceed the most extreme seasonal temperatures recorded in the past century (Battisti and Naylor, 2009).

Together with rice and maize, wheat provides at least 30% of the food calories to more than 4.5 billion people in over 100 countries (Asima et al., 2017). Wheat production is highly sensitive to climatic and environmental variations and is especially sensitive to heat (Porter and Semenov, 2005).

Durum wheat (*Tricitum turgidum subsp. Durum*) is a tetraploid species 2n=2X=28, genome AABB, characterized by bearded spikes with large grain. It is an important crop for the human diet (e.g., pasta, couscous, bourghul, bread, etc.), particularly in the Mediterranean basin where 75% of the world's durum grain is produced, mainly in warm and dry areas (Nachit M.M. 1997). The largest producers in the Mediterranean basin are Syria, Turkey and Italy followed by Morocco, Algeria, Spain, France and Tunisia (Maccaferri et al., 2003; Peng et al., 2011). The major environmental constraints limiting the production in this region are drought and temperature extremes. Hence, the impact of climate change could be significant in terms of raising temperature and reduction of precipitation profiles in the Mediterranean basin. According to the Intergovernmental Panel on Climate Change (IPCC, 2007, 2013) the rise in mean global temperature will be as high as 6.4°C by 2100 with 1.3 times as much CO_2 entering the atmosphere. The report has also predicted a 3-5°C increase in temperature which will result in 20% loss in soil moisture, while the annual precipitation is likely to decrease by 4-27%. Therefore, these changes of seasonality will impose

direct stress not only in the crop performances but also affect the dynamics of pests, diseases and weed.

High temperatures result in a reduction in crop yields by affecting an array of physiological, biochemical and molecular processes (Asima et al., 2017). Heat stress along with drought also showed pronounced effect on grain yield, biomass, days to maturity and grain weight in wheat (Tahir et al., 2006; Kaur and Behl,2010; Balla et al., 2014). In late sown wheat, terminal heat stress is the main cause of yield reduction which is responsible for shortening of grain growth period and improper grain filling (Reynolds et al. 2000; Rane et al., 2007). Every 1 °C rise in temperature above 28 °C during grain filling, results in yield reduction by 3–4 % (Reynolds et al., 1994, 1998; Wardlaw et al., 1989). The most significant factors associated with yield reduction under heat stress are increased sterility, shortened life cycle, reduced light interception and the perturbation of carbon assimilation processes (photosynthesis, transpiration, and respiration) (Reynolds et al., 2010).

Plants respond to high temperature stress by triggering a cascade of events and adapt by switching on numerous stress-responsive genes. However, the complex and poorly understood mechanism of heat tolerance (HT), limited access to the precise phenotyping techniques, and above all, the substantial G × E effects offer major bottlenecks to the progress of breeding for improving HT. Therefore, focus should be given to assess the crop diversity, and targeting the adaptive/morpho-physiological traits while making selections. Equally important is the rapid and precise introgression of the HT-related gene(s)/QTLs to the heat-susceptible cultivars to recover the genotypes with enhanced HT. Therefore, the progressive tailoring of the heat-tolerant genotypes demands a rational integration of molecular breeding, functional genomics and transgenic technologies reinforced with the next-generation phenomics facilities.
As part of the plant's phenotype, yield is the result of the expression of the genotype (G), the environment (E) and their interaction (G x E). In the field of agriculture, management practices (M) are often included as a separate third factor, leading to the G x E x M model. Thus, yield improvement can in principle be achieved by adapting the genotype, the environment, or the management practices. At the level of farming, a few technical and management adjustments may contribute to an increased ability of crops to cope with temperature changes. Firstly, assuming concomitant higher winter temperatures, the dates of planting can be adapted to avoid heat stress later in

the growing season (Olesen and Bindi, 2002; Easterling, 1996; Rosenzweig and Tubiello, 2007; Lotze-Campen and Schellnhuber, 2009). Early planting seems to be a successful strategy to avoid summer heat for maize and spring wheat (Reilly et al. 2003). Secondly, improvements in water management can alleviate heat stress in agriculture, as plants transpire to keep foliage temperature under control. One option is shifting from rain-fed to irrigated agriculture, including low-cost ‘‘rainwater harvesting’’ practices. Additionally, adjusting the timing of irrigation may ensure a crop's water supply at critical, temperature-sensitive stages (Easterling, 1996; Smithers and Blay-Palmer, 2001; Smit and Skinner, 2002; Lotze-Campen and Schellnhuber, 2009). By 2050 and due to climate change, 51% of the region might suffer from a significant reduction in wheat yields unless farmers adopt appropriate cultivars and crop management practices (Ortiz et al., 2008).

Globally, the demand for wheat by the year 2020 is forecasted to be around 950 million tonnes (Singh et al., 2011). This target will be achieved only if global wheat production will be increased by 2.5% per annum (Singh et al., 2011). Poor productivity growth or stagnation in the Green Revolution areas of South Asia and low yields in Africa, coupled with climate change, will make it more difficult to meet the growing demand for wheat (Rosegrant et al., 2009). New wheat cultivars better adapted for future climatic conditions will therefore be required. However, the intrinsic uncertainty of climate change predictions poses a challenge to plant breeders and crop scientists who have limited time and resources and must select the most appropriate traits for improvement (Foulkes et al., 2011; Semenov and Halford, 2009; Zhang et al., 2012).

Breeding for heat tolerance was relied earlier on classical selection methods, however with the advances in understanding biochemical, physiological and molecular basis of heat tolerance, surrogate traits have been identified and are being used in addition to classical methods for selection of heat tolerant genotype. Use of these surrogates in breeding program depends mainly on their correlation with yield under stressed conditions, genetic variability in the available germplasm, heritability and genotype × environment interaction as source of variance for these traits (Kumar et al., 2013).

Based on previous studies, it has been revealed that heat tolerance is under polygenic control in cereals. Different facets of tolerance are determined by different sets of genes which are specific to different growth stages of the plant and various tissues

(Howarth, 2005; Bohnert et al., 2006). Quantitative trait loci analysis is most applicable to the analysis of the genetic basis of thermo-tolerance in cereals.

Aim of the study

Durum wheat as many crops is very sensitive to heat stress, thus breeding for heat tolerant line is becoming very urgent knowing that the warming climate change is leading to increase more and more.

Our study is aimed to:

1- Conduct a molecular assessment of a global durum wheat collection of cultivars, elite breeding lines and landraces, to photograph the current state of germplasm exchange and overall available genetic diversity.

2- The identification of heat susceptibility indexes (HSIs) for several yield-related traits recorded on a panel of modern durum cultivars grown in the South of Morocco and under severe heat along the Senegal River. Therefore, the identification of the loci responsible for guiding heat tolerance to be then deployed in breeding.

3- To prepare for the application of genomic selection (GS) on the heat tolerant traits, four recombinant inbreed lines (RILs) of durum wheat were used to train genomic prediction different statistical models.

Chapters description

Chapter I-Literature review: In this chapter, we gave a description of durum wheat, its origin, evolution and life cycle, the limitations brought by heat stress. Also, we discussed the breeding methods and technologies used in breeding programs.

Chapter II-Genetic Diversity within a Global Panel of Durum Wheat (*Triticum durum*) Landraces and Modern Germplasm Reveals the History of Alleles Exchange: This chapter represents a published paper in Frontiers on 2017, where we assessed a genetic diversity and population structure of durum wheat core collection using different statistical methods: STRUCTURE, DAPC, Neighborhood joining. The obtained results were promising.

Chapter III- Heat Tolerant *Loci* of Durum Wheat Defined via a North-South Gradient: This chapter represents the results of a typical genome wide association mapping study on durum wheat elite lines planted using north south heat gradient, where heat-related QTLs were assessed in mega-environments for yield-related traits and for the Heat susceptibility index of these traits.

Chapter IV-Genomic modeling of major and minor alleles of four durum wheat populations under drought conditions: This chapter summarizes the results of an empirical study of genomic selection on four recombinant inbreed lines planted in five different environments. Several cross validations were made to determine the best possible scenario of genomic prediction in bi-parental durum wheat populations.

Chapter I: literature review

I-1. History and origin of durum wheat

Durum wheat is a monocotyledon from Gramineae family, tribe of *Triticeae* and genus Triticum, which is a classic example of allopolyploidy, its homoeologous genomes are derived from inter-species hybridization belonging to the same family (ozkan et al., 2002). Durum wheat (*Triticum turgidum* L. var. *Durum*) is a tetraploid (2n = 28, AABB) having seven pairs of homoeologous chromosomes with two different genomes A and B. The A genome comes from wild wheat *Triticum urartu Tum*. Better known as *einkorn (Triticum monococcum* L.) with diploid genome AA. While, genome B comes from the wild species, diploid as well, *Aegilops speltoides Tausch* (Figure I-1), Durum wheat (*Triticum durum* L.) is one of the oldest cultivated cereal species in the world. The earliest wheats, dated to approximately 10,000 years BP, were domesticated in the Fertile Crescent, a region extending from the coast of Israel to South-eastern Turkey and westwards through Syria, Iraq and western Iran (Feldman 2001). Durum wheat spread from this region westward into the Mediterranean basin, reaching the Iberian Peninsula around 7,000 years BP (Feldman 2001; MacKey 2005). It is widely accepted that durum wheat entered to North Africa and the Iberian Peninsula from the South of Italy (MacKey 2005). However, recent findings based on the genetic similarities between landraces (LR) from the Maghreb countries and those from Spain and Portugal have suggested North Africa as an additional route for wheat introduction in the Iberian Peninsula (Moragues et al. 2006c, 2007).

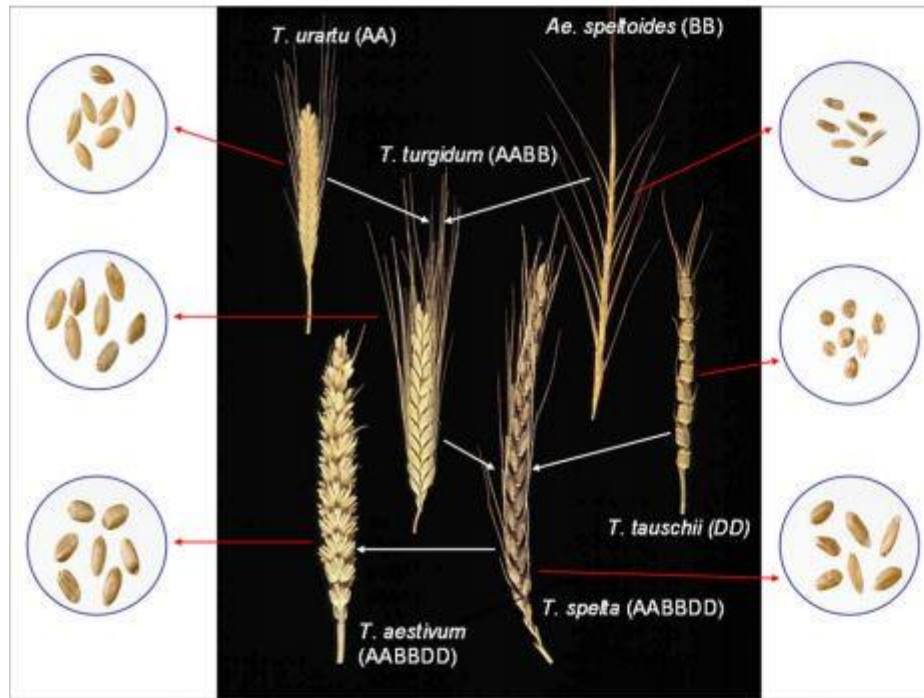

Figure I- 1: The evolutionary and genome relationships between cultivated bread and durum wheats and related wild diploid grasses, showing examples of spikes and grain (Shewry, 2009).

I-2. Durum wheat life cycle

The cycle of development durum includes five phases (Figure I-2):

- ✓ **Germination-emergence**: during germination, the coleorhizae thickens into a white mass and breaks the seed coat of the seed at the level of the germ; this is the beginning of the emission of the primary roots, filled with absorbent hairs. At the same time, the coleoptile, sheathing the first real leaf, elongates towards the surface, where it lets drill the first leaf, it is the lifting. The second and third sheets follow well after.
- ✓ **Tillering**: as soon as the third leaf is emitted, the second internode that carries the terminal bud elongates inside the coleoptile and stops its rise to 2 cm below the surface of the ground, to form the tillering plateau. In the axils of the leaves (from the fourth leaf), axillary buds then enter into activity to give new tillers. The first tiller is formed at the base of the first leaf and the second tiller at the base of the second leaf. The axillary buds in the axils of the leaves of the tillers give rise to the emission of secondary tillers.

- ✓ **Bolting-swelling:** it is characterized by the rise of the spike as a result of elongation between nodes that make up the thatch. The rising tillers enter competitions for the middle factors with the herbaceous tillers which thus do not manage to rise in spikes in their turn. The latter regress and die (Masle, 1982). This phenomenon is manifested in young tillers by a decrease in growth and then a cessation of growth (Masle, 1981).
- ✓ **Heading-Flowering**: once the spike emerging from the sheath of the flag leaf, the heading stage, during which the formation of floral organs ends. Flowering begins 4-5 days later. During flowering, the flowers usually remain closed (cleistogamous flowers), and the three anthers burst and release the pollen (anthesis). Flowers rarely open before pollen release. Flowering lasts from three to six days, depending on weather conditions. It starts at the center of the spike, then continues towards the two extremes of the spike. The receptivity duration of wheat stigma depends on the variety and conditions of the medium, but is between 3 to 13 days. Once fertilized, the ovary grows rapidly. two weeks after fertilization, the embryo is physiologically functional and can produce a new seedling (Bozzini, 1988).
- ✓ **Filling and maturation of the grain**: this is the last phase of the vegetative cycle. It corresponds to the development of the last constitutive component of yield, which is the weight of the grain, following the migration of the carbohydrate substances produced by the flag leaf and stored in the peduncle of the ear (Gate, 1995). It requires heat and dry weather, it will take place in several stages, the milky maturity (the grain still contains 50% moisture and protein storage is coming to an end), the physiological maturity (grain lost in moisture and starch has been formed), complete maturity (moisture content is about 20%), the grain is ripe and ready to harvest, so it is the harvest season.

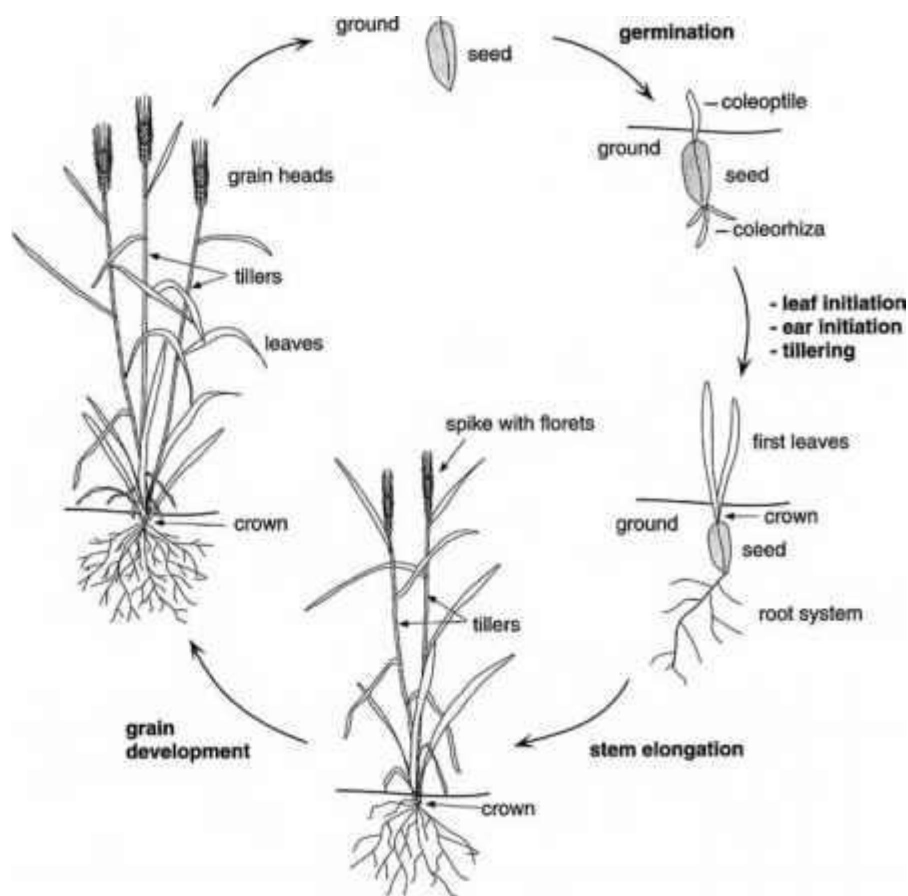

Figure I- 2: Life cycle of wheat. Line drawing by Tim F. Knight. Adapted from E. J. M Kirby and Margaret Appleyard, Cereal Development Guide, 2d ed. (Warwickshire, England: National Agricultural Centre, 1987)

I-3. Cultivation requirements of Durum wheat

Durum wheat is well suited to relatively dry climates where it is warm during the day and cool at night during the growing season, which is typical of Mediterranean and temperate climates. Cultivated in a range of different environments, durum wheat has a very wide adaptability. Seeds can rise below 2 °C even if the optimal temperature is 15 ° C (Bozzini, 1988). Most of the durum wheat produced in the world is spring wheat; however, there are varieties of winter durum (which require vernalization to initiate the transition from vegetative phase to reproductive phase).

In Mediterranean conditions, high temperatures above 30 °C, are stressful. As for low temperatures and cold tolerance, durum wheat has the capacity to withstand temperatures below 4 ° C considered the minimum temperature for growth.

In the Mediterranean region, the water requirements of the crop range from 450 to 650 mm, with relatively low requirements at the beginning of the cycle. It is from the 1 cm ear phase until flowering that they are the most important.

Durum wheat is particularly greedy for Nitrogen to reach a satisfactory level of protein for pasta and semolina manufacturers. Nitrogen intakes should be divided according to the vegetative cycle stages. At tillering, the influence of Nitrogen is manifested on the first component of yield: the number of tillers per plant. In the run-up stage, the Nitrogen supplied makes it possible to emit ears, the number of which is strongly influenced by nitrogen nutrition. Nitrogen deficiency could also result in lower fertility of the ears. At the heading stage, the requirements become very important and the nitrogen demand increases in connection with the growth activity.

I-4. Constraints of durum wheat production

Wheat plants are exposed to many biotic and abiotic stresses that have significant effects on growth and cause changes in the normal physiological functioning of plants (Tas and Tas, 2007, Barnabas et al., 2008). Drought and heat are the most important abiotic stresses affecting plant growth and development (Noohi et al., 2009). It negatively affects wheat growth in many areas of production and is a major factor limiting global wheat productivity (Mohammadi et al., 2004, Reynolds and Trethowan, 2007). Heat stress has many adverse effects on wheat at various stages of growth. High temperatures shorten the tillering phase, resulting in poor establishment of fertile tillers (Baldy, 1984). Heat stress during stem elongation causes deformation and sterility of corn (Combe and Picard, 1994). When stress occurs at the time of flowering, it can reduce pollen vitality and fertility during pollen formation (Barlow et al., 2015). During the grain filling period, heat stress reduces grain size and weight (Dias and Lidon, 2009). For temperatures above 30 ° C, grain weight is systematically reduced (Zahedi and Jenner, 2003). With climate change it is expected a rise in temperatures negatively impacting the productivity of cereals such as durum wheat.

I-5. Breeding methods and technologies

Due to global warming and changes in the climate pattern, development of heat tolerant varieties and generation of improved pre-breeding material seems to be the main focus of any breeding program in future (Ortiz et al. 2007). However, due to complex nature of heat stress tolerance, the advanced techniques of molecular

breeding and genetic engineering in combination with conventional breeding approaches can play a vital role in designing new wheat cultivars with enhanced heat tolerance.

I-5-1. Conventional breeding and plant improvement

The breeding process begins with the production of F1 hybrids by crossing two parents, or more collectively with the most desirable traits for the new variety. In the case of self-pollinating populations such as wheat, the F2 generation derived from the self-fertilization of F1 displays great genetic variability. The selection of the plants having the desired characters begins in F2. It continues until F7-F8, when the offspring becomes uniform. This early selection is done based on characters whose expression will depend little on environmental conditions, such as height, precocity, susceptibility to disease. Selection for complex traits, such as yield and grain quality, is made later on more genetically homogeneous material, using more rigorous devices to control environmental variation (Anonymous, 2006).

I-5-2. Experimental designs

Because all biological material has inherent variability, replicated plots are needed to assess the performance of a variety. To compare the relative performance of several varieties under the same conditions, replicated plots of all the varieties are grown on a contiguous area. These plots are collectively referred to as a variety trial.

In addition to the data from the trial providing information about the relative performance of varieties under the same conditions, the analysis of the data is used to assess the confidence that we have in the results. More confidence can be placed in the results of a trial with a lower level of variation than one with a higher level of variation. Trial designs suitable for variety trials are:

- ✓ **Completely randomized design**

Suppose that 6 varieties were to be compared in a design of 4 replicates.

This can be done by randomly assigning the varieties, A, B, C, D, E and F, to the 24 experimental plots as shown below.

C	A	E	C	B	D
F	C	F	B	F	C
A	D	D	F	E	B
A	E	E	A	D	B

The random allocation of varieties to the plots ensures that each variety has an equal chance of being assigned to each plot. Randomization also protects against source of bias and is also necessary if significance tests are to be made. One disadvantage of the design is that the allocation of varieties to plots may be advantageous to some varieties and not to others. Another disadvantage is that if the trial area is large, the trial could be subjected to positional effects such as fertility gradients or previous cropping which will produce inaccurate variety comparisons. The design will be most useful in a controlled environment experiment or when the area covered by the experiment is small, for example in a growth chamber or in single plant experiments covering a small area.

- ✓ **Complete (or randomized) block design**

If we know, or anticipate that there are or might be positional effects, or even if we know nothing about the trial area, a much sounder experimental procedure is to use blocking. This is a device used to minimize within block variation from sources other than the factors under test and utilizes the fact that plots close to each other are more likely to be similar than those farther apart. The trial area is divided into homogeneous units known as blocks and in a complete block design all varieties are grown once in each of these blocks (replicates). A different randomization is used for each replicate. An example of a complete block design with 6 varieties, A, B, C, D, E and F, in 4 replicates is shown below. Each variety has been randomly assigned to one plot within each replication

Replication 1

C	E	B	D	F	A

Replication 2

F	D	E	B	C	A

Replication 3

E	C	A	F	D	B

Replication 4

D	A	F	E	B	C

Complete block designs are very commonly used in agricultural experimentation due to their simplicity and ease of analysis. Replicate size increases as the number of varieties increases. All plots in a replicate should be contiguous and therefore the larger the replicate the more likely it is that large variation between plots within the replicates will occur. Unless the trial site is known to be very uniform then an experiment containing a large number of varieties (usually 15 varieties or more) might be better conducted using an incomplete block design.

- ✓ **Incomplete Block designs**

When there is a large number of varieties to test, say 25 instead of 6, each replicate of 25 plots is likely to cover a large area of land with an increasing risk of lack of uniformity within the replicate. To insure against such lack of uniformity, further blocking is carried out with each replicate being sub-divided into smaller units known as blocks. It is assumed that there is greater homogeneity within each block than can be expected between the blocks. Such designs are called incomplete block designs. Included in these designs are specific incomplete block designs such as balanced, square and rectangular lattices but these designs are restrictive. The requirements of the statutory and non-statutory variety led to the development of generalized lattice designs. An example of an incomplete block design for 14 varieties (A, B, C, etc.) grown in 2 replicates each containing 3 blocks is shown below.

Replication 1

H	I	E	L	B	K	D	G	N	A	F	M	J	C
2	2	2	2	2	1	1	1	1	1	3	3	3	3

Replication 2

L	A	E	J	C	D	K	I	H	N	B	G	M	F
1	1	1	1	3	3	3	3	3	2	2	2	2	2

For incomplete block designs, the number of plots per block should be chosen so that the area covered by a block is approximately square. With the plot sizes used for the National List and Recommended List cereal variety trials, this results in a block size of between 4 and 8 plots. All plots within a block must be contiguous.

I-5-3. Selection methods in breeding programs

For inbreeding crops, the finished variety is usually a true breeding homozygous line. Five breeding methods are in general use: pedigree selection, single seed descent, doubled haploid production, bulk selection and backcross breeding. Common to all these approaches is the initial requirement to decide which pairs of available parents to cross and enter into the breeding program. There is no generally approved way of deciding what to do here (other than "cross the best with the best") but various attempts have been made to develop methods. Quantitative methods recognize that there is often a compromise between the average merit of the progeny that a cross might produce and the variability among those progenies and try to select parental pairs which maximize the chance of achieving a given target (The National Institute of Agricultural Botany Cambridge, 2017).

- ✓ **Pedigree selection:**

The origins of pedigree selection explicit at the Svalof institute in Sweden, in particular through the work of around 1909 (Åkerberg 1986). Pedigree selection (Figure I-3) is the most common and well-established method of breeding inbreeding crops. It is the gold standard by which other methods are judged. A cross is made between two homozygous parental lines. Selection starts in the F2, the first segregating generation. Selfed seed from the selected individuals are then grown out as single plant progenies, in rows or small plots. Then, over successive generations of selfing, selection occurs between families and between plants within families, with the progeny of the selected plants taken on to the next generation. Selection in these stages is generally on traits which can be scored visually: height, flowering time, disease resistance, but not yield. After several generations of this process, seed is bulked up from each of the final few selected potential new varieties. the inbreeding process and selection occur simultaneously. There is a hierarchical structure to the selection process and some of the lines that finally enter trial may originate from the same selected plant in a generation. In figure 3 the three lines that are produced from this cross (there could be many more) originate from the same F3 row and therefore from the same F2 plant. This is an extreme case. It could be a good thing – if the breeder did little more at the F2 stage than chose plants at random.

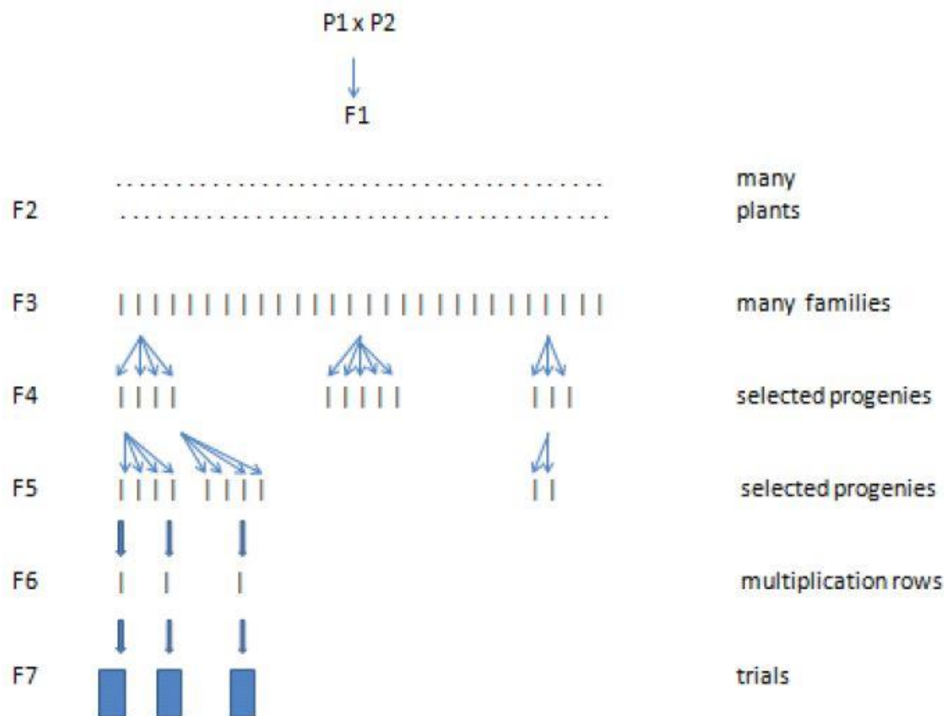

Figure I- 3: Pedigree selection method

- **Single Seed Descent.**

Single seed descent (SSD) (Figure I-4) came of age in the 1070s (Knott and Kumar, 1975) though its conception was earlier (Goulden, 1939). It decouples inbreeding from the selection process. Selfing without selection is carried out over as many generations as are deemed necessary. To maintain variation within the cross, at each generation of selfing only a signle seed is progressed to the next generation from each parental plant (In practice two or three seeds are sown from each plant and are thinned to one, ideally at random.) Each resulting inbred line is therefore derived from a separate F2 individual. In case of little faith in the ability of breeders to select improved lines by pedigree selection and view them as merely throwing away genetic variation at random before proper assessment can occur in yield trial, then single seed descent clearly has advantages. SSD takes several years less to get new varieties into trial compared to pedigree selection. Various compromise schemes also exist, for example the F2 may be grown out in the field and single plants selected, but inbreeding from those single plants is then by SSD

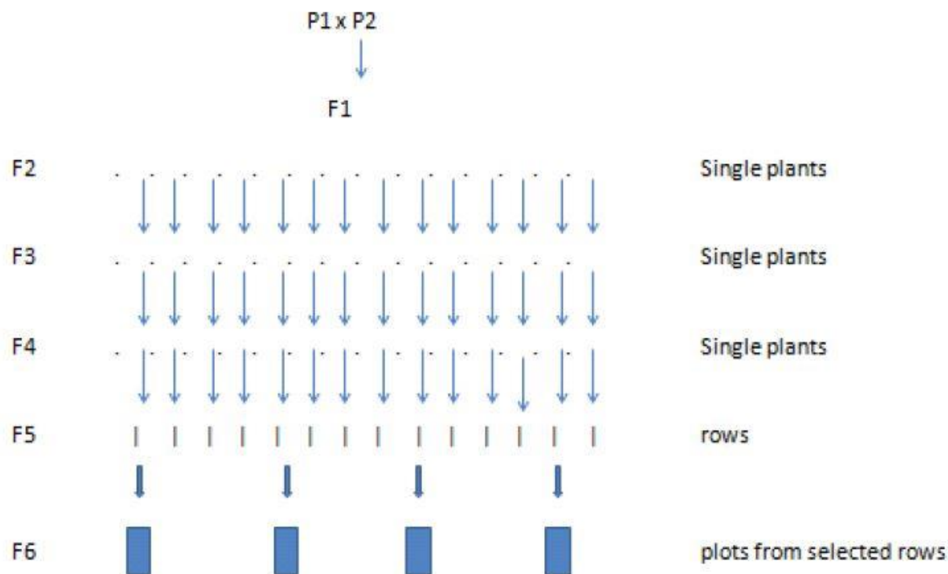

Figure I- 4: Single seed descent (SSD)selection method

- ✓ **Double haploid production**

The role of double haploid (DH) in genetics also has a long history (Imperial Bureau of plant genetics, 1936) but its routine use is much more recent. The basic principle (Figure I-5) is to take gametes from an F1, and culture these somehow to create haploid individuals. These are then treated to double up chromosome numbers and create diploid individuals which are thus completely homozygous. These are multiplied and field tested. This process does not work well in all crops and it is expensive. It is usually the male gametes that are cultured, and the drug colchicine (which interferes with spindle formation during meiosis) is usually used to double up the chromosome numbers. Wide crosses and embryo rescue are involved in some protocols: the chromosomes from the male species are selectively eliminated during mitosis to leave a haploid copy of the genome of the female. One consequence of the expense in creating DH lines is that there is generally less selection at the multiplication stage than with SSD: there is a reluctance to throw out these expensive lines without testing. There is therefore greater care in selecting parents: crosses are less likely to be made speculatively.

The two advantages of DH production are that the lines produced are completely homozygous and that this process is quick. With selfing expected heterozygosity declines at a rate of ½ per generation, so even at F6 1/32th of the genome is heterozygous (on average) and it takes many generations to be confident that you have a completely homozygous line. many breeders are content to test lines in yield trials at

the F3 or F4, in which case the time taken to create new lines is little different between SSD and DH and the merits of complete homozygosity are not viewed to outweigh the costs. As a result, some breeders of crops such as barley, where DH production is well established, prefer to use SSD for routine breeding. DH production may be used for a small number of elite crosses where there is high expectation that high value progeny lines will be found and where there is even greater urgency than usual to deliver. DH progeny lines are less recombined that SSD lines.

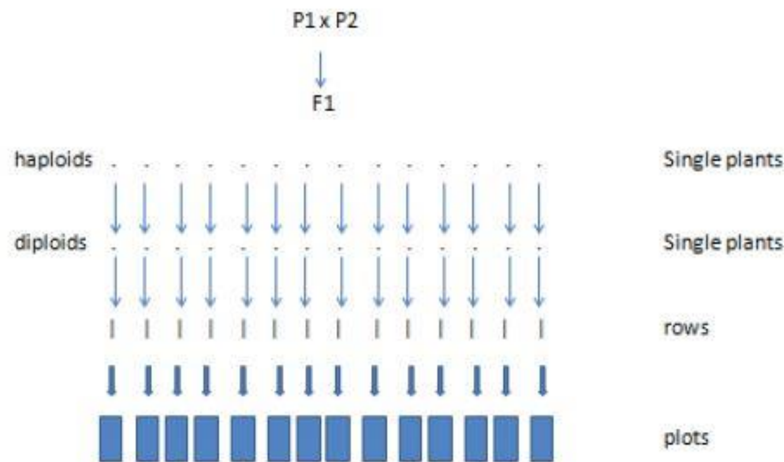

Figure I- 5: Double haploid production

- ✓ **Bulk breeding**

Bulk breeding is another old method, the development of which can also be attributed to Nilssoon-Ehle at Svalof. He wanted a breeding method which would allow him to handle more crosses than he could manage with pedigree selection. The principle (Figure I-6) is that the F2 is allowed to open pollinate in the field (so in most cereals it will predominantly self) and seed is harvested as a bulk (rather than as single plant progenies). The F3 is then grown and harvested in bulk too, and so on over generations. At some generation, say the F4, single plants are harvested separately, multiplied up in small plots or rows, and entered into yield trials. In contrast to pedigree selection, there is no selection of single plants or of individual selfed progenies until the end of the inbreeding process, which is much less effort.

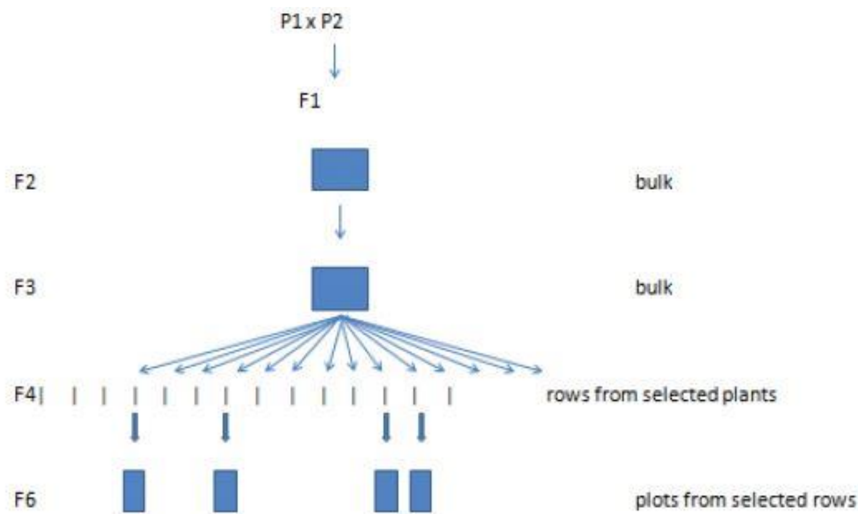

Figure I- 6: Bulk breeding selection method

Much is made of the opportunity for natural selection to act favorably during the bulk inbreeding process: to improve disease resistance for example. However, natural selection can also act against the breeder's interest: favoring taller more competitive genotypes, or favoring genotypes which produce more, but smaller, seed. Consequently, the breeder may select to carry out some selection within the bulks, rouging out very late or early maturing individuals, individuals which are too tall, and sieving seed to maintain seed size. Bulk breeding can also be used under SSD-like conditions to shorten generation time and generate recombinant inbred lines rapidly without the requirement to harvest and sow single seeds in each line of descent. However, more genetic variation will be lost through sampling than in an equivalent SSD program, since many lines may originate from the same F2 (or F3, F4, etc) ancestor, while other F2 individuals will leave no progeny. For these reasons, breeders tend not to favor bulk breeding anymore: it should be less effective than pedigree selection as a means of selecting during inbreeding and less effective than SSD as a means of sampling genetic variation within a cross. For most breeders, the principle merit of bulk inbreeding, that it is cheap, does not outweigh these disadvantages. However, the application of cheap molecular markers could allow selection to maintain genetic variation within bulks and could thus make bulk inbreeding as effective as SSD, but cheaper.

- **Backcross breeding:**

If an elite line is lacking in a particular trait, one may wish to introduce that trait from a second, otherwise inferior line. Backcross breeding makes repeated backcrosses to the line to be improved, accompanied by selection for the missing trait (Figure I-7).

Over a series of backcrosses, the backcross generation comes increasingly to resemble the recurrent parent except for the introgressed trait. Once backcrossing is complete, it is followed by a generation of selfing and selection to create a line which is homozygous at the introgressed trait loci. Backcross breeding has been used primarily to introgress major genes for disease resistance from unadapted germplasm into elite material. In practice, backcrossing will be stopped as early as the breeder dares, to save time. If the non-recurrent parent is highly unadapted, more generations of backcrossing will be required. Typically, a minimum of two backcrosses are carried out. If the trait to be introgressed is a dominant major gene, selection in each the backcross generation is simple, since heterozygotes for the gene can be identified. If the trait is recessive, then large backcross populations must be carried forward so that on selfing there is a high probability that homozygotes for the trait locus will be found. Alternatively, backcrossed individuals must also be selfed and the progeny scored to confirm that they are carrying the recessive allele. Various modifications of this procedure as possible, but in recent decades' marker assisted breeding has had a big impact since the trait locus can be tagged and identified in heterozygous form.

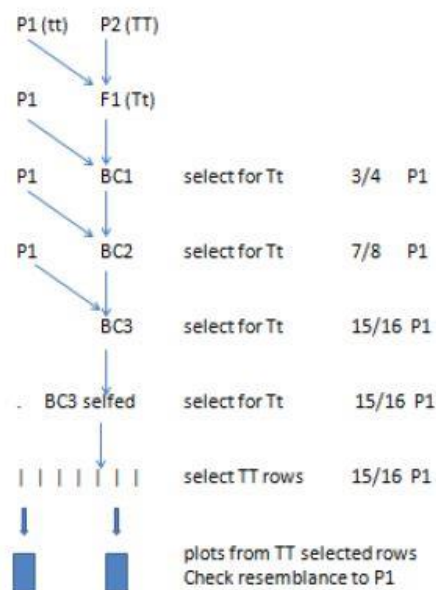

Figure I- 7: Backcross selection method

Backcross breeding need not be restricted to introgression of single loci, though the greater the number of loci to be introgressed simultaneously, the more complex the procedure becomes. In an extreme form, there are no known loci to be introgressed and pedigree breeding or SSD is initiated in the first or second backcross rather than in the F2. Genetically, this can sometimes be the right thing to do; if your best line is much better than your second best. Sociologically it can sometimes be the right thing to do too; farmers can be reluctant to change from an established variety they understand and can grow well. In these circumstances it can do better to release an "improved" version of an existing variety rather than a new variety with a completely different look and feel.

I-5-4. Genome wide association mapping (GWAS)

As the fundamental aim of genetics is to connect genotype to phenotype (Botstein and Risch, 2003), association mapping seeks to identify specific functional variants (i.e., loci, alleles) linked to phenotypic differences in a trait, to facilitate detection of trait causing DNA sequence polymorphisms and/or selection of genotypes that closely resemble the phenotype (Wang et al, 2005). Association mapping has been variously defined (Chakraborty and Weiss, 1988; Kruglyak, 1999; Zhu et al., 2008), and has also been referred to as "association genetics", "association studies," and "linkage disequilibrium mapping" although the latter term is also used to reflect studies detecting associations among loci. The general characteristics of this field of genetics involve the use of unstructured or loosely structured populations – usually intraspecific – that are both phenotypically and genotypically characterized to detect statistical associations between genetic polymorphisms and heritable trait variation. Some experimental designs involve use of progenies. The actual polymorphisms causing trait variation are usually not known, and therefore are not directly observed, but rather, are detected via statistical inference. The predicating condition for detection of such associations is nonrandom association of causative trait polymorphisms with observed polymorphisms, i.e., linkage disequilibrium (LD). Association genetics is a multidisciplinary field, involving components of genomics, statistical genetics, molecular biology, and bioinformatics which together form the basis for selecting, evaluating, and associating genomic regions for correlation with trait variation (Figure I-8). The higher resolution afforded by use of unstructured populations allows the intriguing possibility of identifying the genes or even the specific nucleotides

underpinning trait variation. Also, the opportunity to use molecular markers to enhance rates of genetic gain, including the utilization of specific genes from non-elite germplasm in a more directed and efficient manner than was hitherto possible.

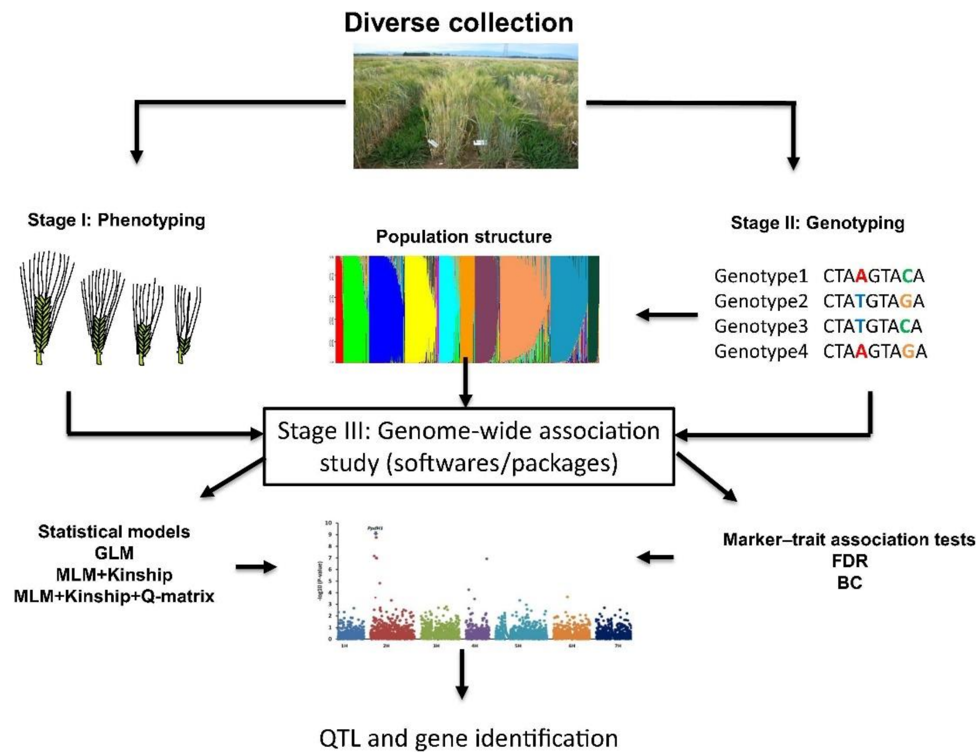

Figure I- 8: Genome wide association mapping principle. Stage I: Phenotyping, stage II: Genotyping and stage III: Genome-wide association study including statistical models, multiple-testing analyses, and software/packages for QTL and gene identification (Alqudah et al., 2020).

Figure I-8 explains GWAS in three major steps (Alqudah et al., 2020). Stage I is the phenotyping in which all genotypes should be phenotyped for a particular trait or group of traits based on the objectives of the study. Stage II is the genotyping in which the same set of individuals that were phenotyped should be used for genotyping using DNA molecular markers. Before running GWAS, population structure should be tested in order to select the better GWAS model. The general linear model (GLM) and mixed linear model (MLM) are statistical models often proposed for performing GWAS. The GLM does not take the population structure-related into account. The MLM, on the other hand, considers the population structure in its model (Kinship or kinship + Q matrix + PCAs).

Finally, the phenotypic and genotypic data are combined using appropriate software (e.g TASSEL: Trait Analysis by Association, Evolution, and Linkage) by which alleles associated with a particular trait can be detected after the GWAS model was selected

(Stage III). The significance of marker-trait associated (passing the threshold e.g. –log10 p-value 3) is usually determined by the false discovery rate (FDR) or Bonferroni correction (BC) which can be defined as multiple comparisons that can be fit to test the significance of hundreds of thousands to millions of markers in GWAS. For BC, the significant level is divided by the number of tests (markers) at each locus. The false discovery rate (FDR) is another test that provides an estimate of the number of actual true results among those called significant (Storey, J. D., & Tibshirani, R., 2003).

I-5-5. Quantitative trait locus (QTL) as a tool to detect genetic variation

Since the advent of complete genome sequences, new, much more favorable perspectives have opened up to identify the genes involved in the genetic variability of phenotypic traits. These sequences have allowed the identification of many genetic markers and recent technological advances now offer the possibility of using high throughput genotyping tools at a relatively low cost. The use of SNP-type markers, present in large numbers, makes it possible to glimpse a more precise location of the gene (s) responsible for a phenotype. Two main approaches are possible in the search for causal genes, the candidate gene approach and QTL mapping. The candidate gene approach requires a priori knowledge on the implication of certain genes on the studied character, it consists in studying the association between a phenotype and alleles of these candidate genes. The QTL mapping, much more used, explores, without prior knowledge, all or part of the genome by testing possible associations between the polymorphism of markers, such as SNPs, and the variability of the phenotype. This approach makes it possible to locate chromosomal regions (QTL) carrying the gene (s) whose polymorphism is involved in the variability of the studied trait. Before the advent of high-throughput SNP chips (between 50K and 1000K), QTLs were detected and localized using biparental populations and microsatellite markers (SSRs) distributed throughout the genome with significant marker distances greater than 20 mega base (Mb). For this type of data, the method of choice is the linkage analysis which consists in looking for a possible intra-family association between the grandparental origin of the chromosomal regions transmitted by the parents and the phenotypes of the descendants. The grandparent origin is traced by markers whose apparent effect can vary from one family to another. The greater the difference between the average phenotypes of the offspring who received the paternal or maternal haplotype, the greater the QTL is important. Nevertheless, since the grandparental origin of the

chromosomal segments received by a descendant is most often identical, the location of the QTL is not very precise. To obtain a more precise localization, it is necessary that recombination (exchange of portions of homologous chromosomes during meiosis) occur during the constitution of the gametes. The more recombination in a region of a given size, the more precise the location but the larger the families. SNP markers, which are distributed over the entire genome at an average interval of about 40-50 Kb (for chips of about 50,000 markers), make it possible to locate QTLs more precisely by using association analyzes. These analyzes exploit the phenomenon of linkage disequilibrium, which is a nonrandom association of alleles at several loci in the population of chromosomes (or gametes) present in the population. The closer two loci are, the higher the association. In contrast, two distant loci (especially on two distinct chromosomes) tend to be in linkage equilibrium. Association studies between a marker (or a set of markers) and a phenotype therefore assume that this marker (or this set) is in linkage disequilibrium with a QTL whose polymorphism causes part of the variability of the character. Several factors such as genetic drift, migration, mutations and selection are responsible for linkage disequilibrium and its evolution in a population is a function of the rate of recombination between loci.

There are several methods for detecting QTL based on the use of linkage disequilibrium, commonly referred to as association methods. The purpose of these methods is to test the association between a marker (or several) and a phenotype in a population of supposedly unrelated individuals. With the large number of SNP markers that cover the entire genome and the fact that they are close enough, we can exploit the linkage imbalance that exists between these markers and between these markers and potential QTLs. Therefore, when an association analysis declares a significant marker, the underlying assumption is that there is a QTL, close to the marker, which is in linkage disequilibrium with it.

I-5-6. Marker assisted selection

Marker assisted selection (MAS) is a type of indirect selection based on a significant association between a marker and variation for target trait. it has been used in plant improvement programs since the 1990s, after promising research results for tagging genes or mapping QTL. MAS and association genetics have been used in the detection of underlying major genes in gene pools and in their introgression to improve traits of major crop breeding programs. The methods of marker-assisted selection (MAS) or

marker-assisted recurrent selection (MARS) assume that the user knows which alleles are favorable, and what their average effects on the phenotype are (Charmet et al., 1999; Hospital et al., 2000; Bernardo and Charcosset, 2006). This assumption is viable for major-gene traits but not for quantitative traits that are influenced by many loci of small effect and the environment. In locus identification and effect estimation for such traits, much uncertainty will remain (Beavis, 1994; Melchinger et al., 1998; Schon et al., 2004).

I-5-7. Genomic selection

Genomic selection or genome-wide selection (GS) has been highlighted as a new approach for marker-assisted selection (MAS) in recent years. The term 'GS' was first introduced by Haley and Visscher at the 6th World Congress on Genetics Applied to Livestock Production at Armidale, Australia in 1998 according to Meuwissen (2007).

GS is a form of MAS that selects favorable individuals based on genomic estimated breeding values (GEBVs). Breeding values have not been a popular index in plant breeding, although they are frequently used in animal breeding. They are defined as 'the sum of the estimate of genetic deviation and the weighted sum of estimates of breed effects' (Van Vleck et al., 1992), which are predicted using phenotypic data from family pedigrees based on the additive infinitesimal model (Fisher, 1918).

The general processes of GS and traditional MAS used for quantitative traits (QTs) are shown in figure I-9. The main frameworks of the two approaches are similar, where both GS and traditional MAS consist of training and breeding phases. In the training phase, phenotypes and genome-wide (GW) genotypes are investigated in a subset of a population, i.e. the training population in GS and the mapping population in traditional MAS. Within populations, significant relationships between phenotypes and genotypes are predicted using statistical approaches. In the breeding phase, genotype data are obtained in a breeding population, before favorable individuals are selected based on the genotype data obtained. Three obvious differences between the two approaches are apparent: (1) in the training phase, quantitative trait loci (QTLs) are identified in traditional MAS while formulae for GEBV prediction are generated in GS, known as GS models; (2) in the breeding phase, genotype data are only required for targeted regions in traditional MAS, whereas GW genotype data are considered to be necessary in GS; (3) in the breeding phase, favorable individuals are selected based on the genotypes of markers in MAS, whereas GEBVs are used for selection in GS. Thus, GS jointly

analyses all the genetic variance of each individual by summing the marker effects of GEBV (Heffner et al., 2009), and it is expected to address small effect genes that cannot be captured by traditional MAS (Hayes et al., 2009)

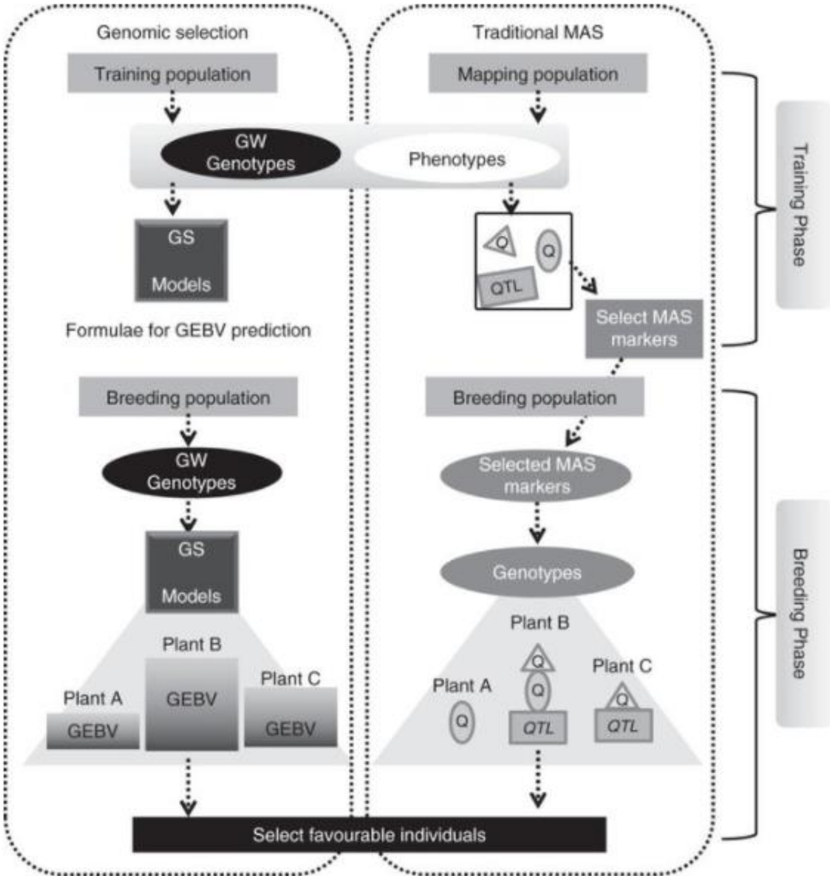

Figure I- 9: Schemes of genomic selection (GS) (left) and traditional MAS for the selection of quantitative traits (right) (Nakaya & Isobe, 2012).

Chapter II: Genetic Diversity within a Global Panel of Durum Wheat (*Triticum durum*) Landraces and Modern Germplasm Reveals the History of Alleles Exchange

Hafssa Kabbaj[a,b], Amadou Tidiane Sall[a,b], Ayed Al-Abdallat[c], Mulatu Geleta[d], Ahmed Amri[b], Abdelkarim Filali-Maltouf[a], Bouchra Belkadi[a], Rodomiro Ortiz[d], and Filippo Maria Bassi[b]

a Laboratory of Microbiology and Molecular Biology, Mohammed V University, Rabat, Morocco
b International Center for the Agricultural Research in the Dry Areas (ICARDA), Rabat, Morocco
c Department of Horticulture and Crop Science, Faculty of Agriculture, The University of Jordan, 11942 Amman, Jordan
d Swedish University of Agricultural Sciences (SLU), Department of Plant Breeding, Sundsvagen 14 Box 101, SE 23053 Alnarp, Sweden

Abstract

Durum wheat is the 10th most important crop in the world, and its use traces back to the origin of agriculture. Unfortunately, in the last century only part of the genetic diversity available for this species has been captured in modern varieties through breeding. Here, the population structure and genetic diversity shared among elites and landraces collected from 32 countries was investigated. A total of 370 entries were genotyped with Axiom 35K array to identify 8,173 segregating single nucleotide polymorphisms (SNPs). Of these, 500 were selected as highly informative with a PIC value above 0.32 and used to test population structure via DAPC, STRUCTURE, and neighbor joining tree. A total of 10 sub-populations could be identified, six constituted by modern cultivars and four by landraces of different geographical origin. Interestingly, genomic comparison among groups indicated that Middle East and Ethiopia had the lowest level of allelic diversity, while breeding programs and landraces collected outside these regions were the richest in rare alleles. Further, phylogenetic analysis among landraces indicated that Ethiopia appears as a second center of origin of durum wheat, rather than a second domestication site as previously believed.

II-1. Introduction

Durum wheat (*Triticum turgidum* ssp. *durum* Desf., $2n = 4x = 28$, AABB) is the 10th most important crop worldwide owing to its annual production of 37 million tons (LMC international, 2009). It is grown on about 10% of the world's wheat area mostly in West Asia, North and East Africa, the North American Great Plains, India, Eastern and Mediterranean Europe (Cantrell, 1987; International Wheat Council, 1991). With the exception of Europe, North Africa (Algeria, Morocco, Tunisia and Libya) is the largest import market for durum wheat (Bonjean et al., 2016). Its final uses vary between industrial production of pasta, couscous, and other semolina products and traditional handmade foods such as *frike*, bourghul, and unleavened breads. The vast array of homemade foods derived from durum grains is the result of its long history as part of human diets, which dates back to the origin of civilization in the Fertile Crescent (Mackey, 2005). Tetraploid wheat domestication took place about 12,000 years ago in the Fertile Crescent, when primitive farmers selected among cultivated forms of wild emmer (*Triticum turgidum* ssp. *dicoccoides*) a naked form that was easier to thresh (*Triticum turgidum* ssp. *dicoccum*; MacKey, 2005; Tanno and Willcox, 2006; Zohary et al., 2012). Approximately 2,000 years after this event, human migration and the spread of agriculture from the Fertile Crescent to and throughout Europe and Asia led to the expansion of the cultivation of naked emmer. During the same period, durum wheat (*Triticum turgidum* ssp. *durum*) appeared in the Fertile Crescent as result of further selection and domestication of naked emmer (Zohary et al., 2012). Due to its larger grains and higher productivity, durum gradually replaced its ancestor to become by the second millennium BC the major cultivated form of tetraploid wheat (Maier, 1996; Nesbitt and Samuel, 1998; Zohary et al., 2012). Thus, durum wheat origin is the result of two successful domestication events by ancient farmers, first from wild emmer to domesticated emmer, and second from cultivated naked forms of emmer to durum (Gioia et al., 2015). The Levant (Jordan, Lebanon, Israel, Palestine and Syria) is considered to be the center of origin of this crop (Vavilov, 1951; Feldman, 2001). From there, it spread throughout the Mediterranean basin, probably via trading by Phoenician merchants, by the caravans' routes along the Sahara Desert or the North African coasts (Bozzini, 1988), and the Silk Road to Asia (Waugh, 2010). Reports (Dejene et al., 2015 and 2016) suggested, that durum wheat was also domesticated a third time to derive *Triticum aethiopicum* Jakubz. (syn. *Tritium durum* subsp. *abyssinicum* Vavilov), which is mainly found today under cultivation in Ethiopia and

neighboring countries. It remains yet unclear if this additional domestication was the result of further modification by farmers of a durum landrace population originated in the Levantine, or rather if it represented a novel origin of durum by domestication of naked emmer. What is clear is that the *abyssinicum* subspecies is morphologically very different, with uncompact spikes and small dark seeds (Sakamoto and Fukui, 1972; Porceddu et al., 1973; Pecetti et al., 1992; Dejene et al., 2015).

The history of durum wheat trading became more complex at the beginning of the 20th century when breeders started imposing selection pressure for commercial purposes (Autrique et al., 1996; Pecetti and Annicchiarico, 1998). In 1910, Nazareno Strampelli set up the first durum wheat breeding program in Foggia, southern Italy. This program was initially based on the selection of pure lines from local landraces (Scarascia, 2005). Later, Strampelli recognized the great value of the inheritance laws described by Mendel and started a true hybridization program. The most successful result was the cultivar 'Cappelli', released in 1915 (Laidò et al., 2013). This pioneer cultivar had a major global impact in the years that followed, and most of the modern varieties can be traced back to the 'Capelli' lineage. A second major impact was provided by the shuttle breeding system developed by the Nobel Laureate Norman E. Borlaug several years later in Mexico and his deployment of dwarfing genes to increase harvest index (Gale and Youssefian, 1985). This resulted in the release of several semi-dwarf and widely adapted cultivars that are still grown nowadays (Ortiz et al., 2007). The modern scenario of the pedigrees post Green Revolution is extremely hard to describe, with several hybridization occurring between different breeding programs and mega-cultivars that have crossed the boundaries of their country of origin. To disentangle the last 40 years of germplasm exchange and cross hybridizations, new methods have been devised based on the allelic similarities described by molecular markers (Christiansen et al. 2002; Pritchard and Rosenberg, 1999; Falush et al., 2003; Flint-Garcia et al., 2003). A bi-product of these type of studies is the understanding of how much of the overall available alleles (namely: genetic diversity) have been captured within a specific germplasm. Since genetic diversity is often seen as an essential source of novel and useful alleles to be selected by breeders (Tanksley and McCouch, 1997; Cooper et al., 2001; Spillane and Gepts, 2001; Acosta-Gallegos et al., 2007), this type of studies have both a historical value and an immediate practical impact on breeding. Hence, the aim of this research was to conduct a molecular assessment of a global durum wheat collection of cultivars, elite breeding lines and landraces, in order

to photograph the current state of germplasm exchange and overall available genetic diversity.

II-2. Materials and Methods

II-2-1. Plant material

A large durum wheat germplasm collection exceeding 1,500 accessions was assembled at the field station of the International Center for Agricultural Research in the Dry Areas (ICARDA) in Terbol, Lebanon (33° 49' 05" N, 35° 58' 59" E). A core subset was defined after assessing the collection for similarity in flowering time, response to toxic level of boron, disease response, tendency to lodge, visual selection, and characterized with 10 single nucleotide polymorphisms (SNPs) associated to known genes. The original set contained several landraces selected on the basis of the algorithm for Focus Identification of Germplasm Sources (FIGS; Mackay et al., 2005; Bari et al., 2012) targeting the model to identify sources of rust resistance, and tolerance to drought, heat and mineral toxicity. A core subset of 384 accessions was selected to be similar in phenology and diverse for all other traits. It includes 96 landraces from 24 countries and 288 cultivars and elite breeding lines from 8 countries, including several breeding lines from both ICARDA and CIMMYT. This panel was built on the work already carried on by Maccaferri et al. (2003), removing duplicates and adding a set of landraces, new breeding material from ICARDA, International Maize and Wheat Improvement Center (CIMMYT), Canada, and Australia (Figure II-1). A detailed list of materials is provided in Annex 1.

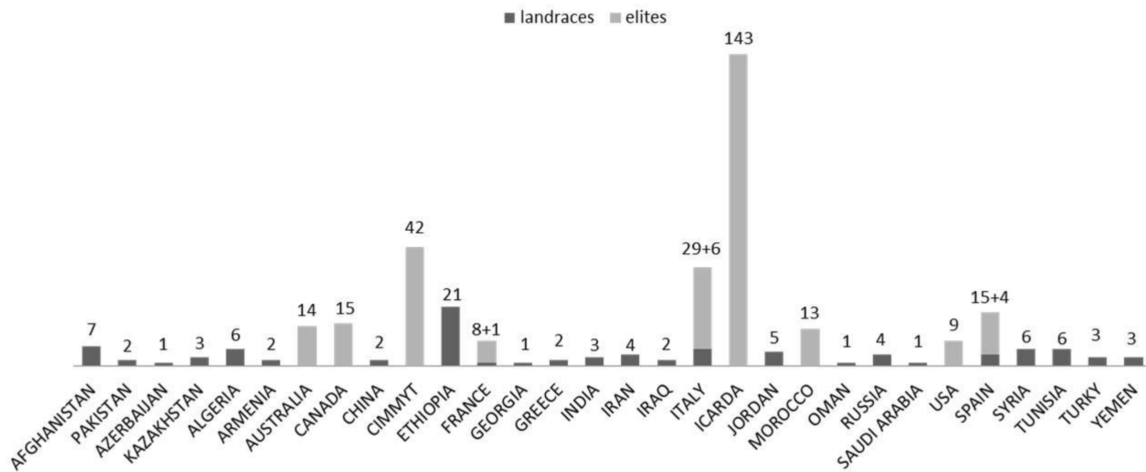

Figure II- 1: Country of origin of the durum wheat core collection lines (landraces in black, elite lines in grey).

II-2-2. DNA extraction and genotyping

DNA was extracted from leaf samples using a standard cetyltrimethylammonium bromide (CTAB) protocol. The 384 accessions were genotyped by 35K Affymetrix Axiom wheat breeders array (www.affymetrix.com) at Trait Genetics (Gatersleben, Germany) following the manufacturer instructions.

II-2-3. Data analysis

The polymorphic information content (PIC) was calculated following the formula described by Botstein et al. (1980), and the 2 points LOD was generated using Carthagene software (Schiex et al., 2009). The discriminant analysis of principal components (DAPC), was performed using the 'adegenet' package 1.4-1 (Jombart et al., 2010) in R Studio V 2.3.2 (R development Core Team). With DAPC, the genetic variationy was decomposed using a multivariate ANOVA model as:

Total variance VAR(X) = variance between groups B(X) + variance within groups W(X). Other approaches such as Principal Component Analysis (PCA) or Principal Coordinates Analysis (PCoA / MDS) focus on VAR(X). That is, they only describe the global diversity, possibly overlooking differences between groups. On the contrary, DAPC optimizes B(X) while minimizing W(X) (Jombart et al., 2015). The variance explained by PCA was fixed to 75 and the value of *k* was tested from 2 to 50. The rate of decrease of the Bayesian information criterion (BIC) was visually examined (Figure II-1), and the number of clusters was determined as the value of *k* above which BIC values decreased. Analysis of admixture by kinship was performed using the Bayesian clustering algorithm implemented in the software STRUCTURE v 2.3.4 (Pritchard et

al., 2000) using 50,000 burning periods and 10,000 replicates and re-assessed five times with 11 independent runs. The value of k was set based on DAPC results. To further confirm cluster analysis, unweighted pair group method of association (UPGMA) was carried out using the genetic similarity matrix by numerical taxonomy and multivariate analysis system (NTSYS-PC) version 2.02e software (Rohlf 1997). Because this method uses genetic similarity matrix, a line of reference was arbitrary set to explain 60% of similarity in order to determine the genetically distinct branches of the tree. Arlequin 3.5.2.2 (Excoffier et al., 2005) was used to assess the molecular variance (AMOVA) between clusters. Phylogenetic studies of landraces were conducted by neighbor-joining algorithm of the genetic distances determined by STRUCTURE using 1,000 bootstrapping analysis for an unrooted tree by DARwin V 6.0.12 software (Perrier et al., 2003). DIVA-GIS V 7.5.0 software (Hijmans et al., 2012) was used to graphically map the GPS coordinates of the places of collection of the landraces.

II-3. Results

II-3-1. Genotyping of a global panel of durum elites, cultivars and landraces

The 35k Axiom Breeds Array was used for genotyping. This array was developed by choosing tags of proven high polymorphism when tested on modern bread wheat elites, among the 817k SNP Axiom HD platform. A total of 384 durum entries were genotyped, but only 370 showed DNA quality sufficient for SNP calls. In total, 35,143 SNPs were assessed, of these 11,642 (34%) failed to meet the minimum call rate, which suggests that these markers were probably located on the D genome, present in hexaploid bread wheat but not in tetraploid durum wheat. A total of 14,851 (42%) met the quality cutoff but remained monomorphic in this population, while 8,173 (36%) were found to be high quality and polymorphic. The average frequency of the minor allele was 12% with a minimum of 3%. Table II-1 gives the chromosome assignment of markers based on the work by Winfield et al. (2015) in bread wheat, where 1,559 markers remained unassigned.

Table II- 1: Number and distribution across the 14 chromosomes (Chr.) of durum wheat of polymorphic SNPs markers on the Axiom 35K breeder's array and the 2-points LOD for the subset of the 500 SNPs used for clustering.

Chr.[a]	Polymorphic	Subset of 500 SNPs	2-points LOD of the subset		
			Average	MIN	MAX
1A	505	26	36.7	0.3	101.2
1B	617	48	33.9	0	111.1
2A	519	31	36.8	0	111.1
2B	589	28	23.7	0	110.8
3A	411	22	36.8	0	110.5
3B	533	41	30.8	0.1	111.4
4A	306	25	35.8	0	101.5
4B	283	22	39.6	0	110.8
5A	489	33	35.7	0	110.2
5B	673	48	31.4	0	110.2
6A	360	22	49.4	0	111.1
6B	480	40	26.1	0	111.1
7A	505	35	36.5	0	110.5
7B	344	41	41.4	0	111.4
Unassigned	1,559	38	8.5	0	78
Total	8,173	500			

[a] Chromosome assignment was done on the basis of a consensus bread wheat genetic map (Winfield et al., 2015)

II-3-2. Population stratification

A subset of 500 highly polymorphic (0.32≤ PIC ≥0.45) markers was chosen for clustering and kinship studies. These markers were selected for even distribution across the genome, covering all durum chromosomes, with LOD values that ranged from 0 to 111.1, and averaged at a minimum of 23.7 in chromosome 2B and 49.4 in chromosome 6A (Table II-1). These LOD scores indicate good distribution and correct chromosome assignment. DAPC inferred the optimum number of sub-populations to be 10 (Figure II-2). AMOVA was used to determine that variation among and within groups was highly significant (P < 0.001), with the clusters capturing 31.5% of the variations, while 68.3% was explained by individuals within populations (Table II-2).

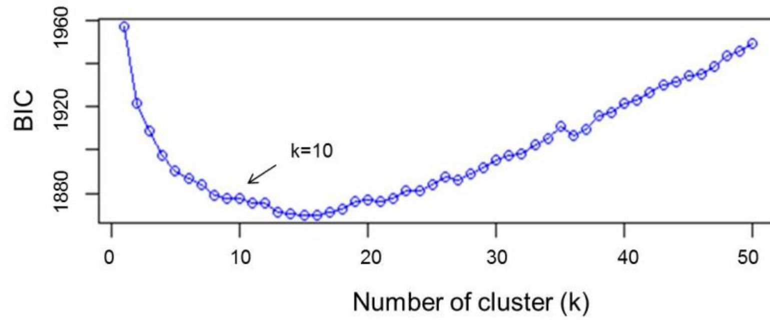

Figure II- 2: Statistical determination of the optimum number of clusters by discriminant analysis of principal components (DAPC).

Table II- 2: Analysis of molecular variance (AMOVA) for stratification of a global durum wheat panel into 10 sub-populations.

SOV	Degrees of freedom	Sum of squares	Variation (%)
Populations (P)	9	20906***	0.315
Individuals within P	360	49809***	0.683
Individuals	370	67***	0.002
Total	739	70783	

*** indicate that source of variation was highly significant at $P \leq 0.001$

Among the 10 clusters, four groups were composed of landraces, while six groups included mostly cultivars and elite lines. STRUCTURE was also used to determine cluster assignment, with the strongest contradiction between DAPC and STRUCTURE identified among landraces. In fact, only two clusters were identified by the latter compared to the four by DAPC. This issue was circumvented by running separately landraces and modern lines, in which case good agreement could be found between the two software. Instead, tree-based studies by UPGMA identified six clusters determined at 60% of similarities, three composed of mostly landraces and three by modern lines. This value was kept as identified. Entries were assigned to cluster based on DAPC study as it was considered the most reliable method, but a qualitative score was given to each assignment as 'solid' when the three methods agreed in the assignment, 'good' when two methods agreed, and 'bad' when only DAPC made the

assignment. Among 370 lines, 16 (4.3 %) were scored as 'bad', 70 (18.9 %) as 'good', and 284 (76.7%) as 'solid' (Table II-3).

Table II- 3: Reliability of the entries assignment based on comparison between three genetic clustering methods: DAPC, STRUCTURE, and neighbor-joining (details available in Table S1).

DAPC Clusters	**Assigned entries**	**Reliability score**		
ID	N	Bad	Good	Solid
1. Middle East	11	0	0	11
2. *T. abyssinicum* type	18	0	0	18
3. Mediterranean trades	26	7	8	11
4. Central and South Asian	27	1	4	22
5. 'Om Rabi' derivatives	13	0	1	12
6. Italian cultivars	26	6	5	15
7. Breeding program exchange	58	2	29	27
8. Developed countries	30	0	4	26
9. ICARDA derived	119	0	5	114
10. CIMMYT derived	42	0	14	28
Total	370	16	70	284

In particular, sub-populations 3 and 6 had the highest number of badly assigned entries. Clusters 1, 2, 5, 8, and 9 were the most solid with no 'bad' assignments, and few 'good'. Full details for each genotype are provided in Annex 1.

Cluster 1 comprises 11 landraces from West Asia (Levantine). Cluster 2 is represented by 18 landraces, 15 from Ethiopia, one from Yemen, one from Jordan and one from Russia. Cluster 3 is composed of 26 landraces, five from Tunisia, four from Algeria and Spain, two from Afghanistan, Greece, and Italy, and one each from Azerbaijan, China, Ethiopia, Iran, Kazkhstan, and Russia. Cluster 4 is composed mainly of Central and South Asian landraces and includes those collected in Afghanistan, Armenia, Georgia, India, Iran, Iraq, Kazakhstan, Pakistan, Turkey, Russia, but also Italy, Oman, Yemen, and Saudi Arabia. Cluster 5 gathers 13 modern lines from the breeding program of ICARDA, which include in their pedigree 'Om Rabi' – a line derived from the cross between the elite 'Jori' and the Jordanian landrace 'Haurani', one Italian landrace and the Italian cultivars 'Arcangelo', 'Appio', and 'Capeiti'. Cluster 6 contains 20 modern lines and 6 landraces, with 13 cultivars and two landraces from Italy, four landraces

from Ethiopia, and the remaining modern germplasm from France, ICARDA and Spain. Cluster 7 is represented by 58 entries from different breeding programs, including 24 elites from ICARDA, six from CIMMYT, four cultivars from France and USA, seven from Italy, six from Morocco and from Spain, but also one landrace each from Spain and France. Cluster 8 includes 17 cultivars from North America, five from Australia, two from France, one from Italy and Spain, plus two landraces from Algeria and one landrace selection ('Shabha') from ICARDA. Cluster 9 is the largest with 106 breeding lines from ICARDA, two from CIMMYT, four varieties from Italy, four from Morocco and Tunisia, 'Wallaroi' from Australia, and two Moroccan landraces. Cluster 10 groups 24 elite lines derived from the breeding program of CIMMYT, one from ICARDA, nine Australian cultivars, five Spanish, two Moroccan, and one Iranian landrace. The clustering of the panel is presented in Figure II-3, and modern lines are detailed in Figure II-3(B).

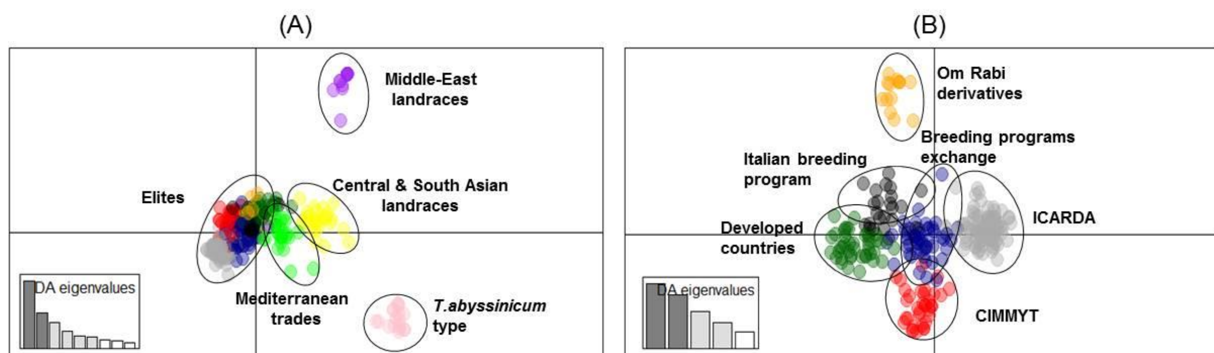

Figure II- 3: Cluster analysis using discriminant analysis of principal components (DAPC). (A) Graphical representation of principal component 1 (IPC1) and 2 (IPC2) distances for 10 sub-populations within the whole panel. (B) Graphical representation of IPC1 and IPC

II-3-3. Admixture analysis by kinship

Admixture analysis was conducted using STRUCTURE. However, this software could not distinguish any subgroups among landraces, when these were analyzed together with modern cultivars (Figure II-4). To better detail the kinship among landraces, these were also analyzed alone in the form of a phylogenetic analysis (Figure II-5). Four main branches could be identified as defined by clusters 1, 2, 3, and a part of 4.

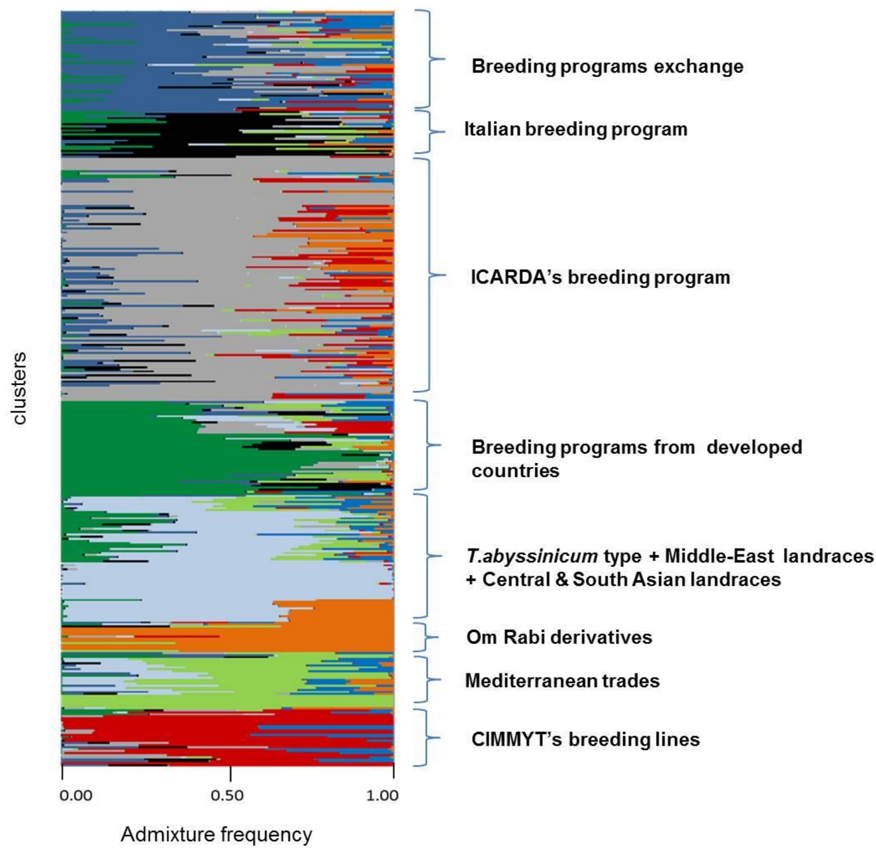

Figure II-4: Admixture kniship matrix for landraces and elite lines of 370 durum wheat individual

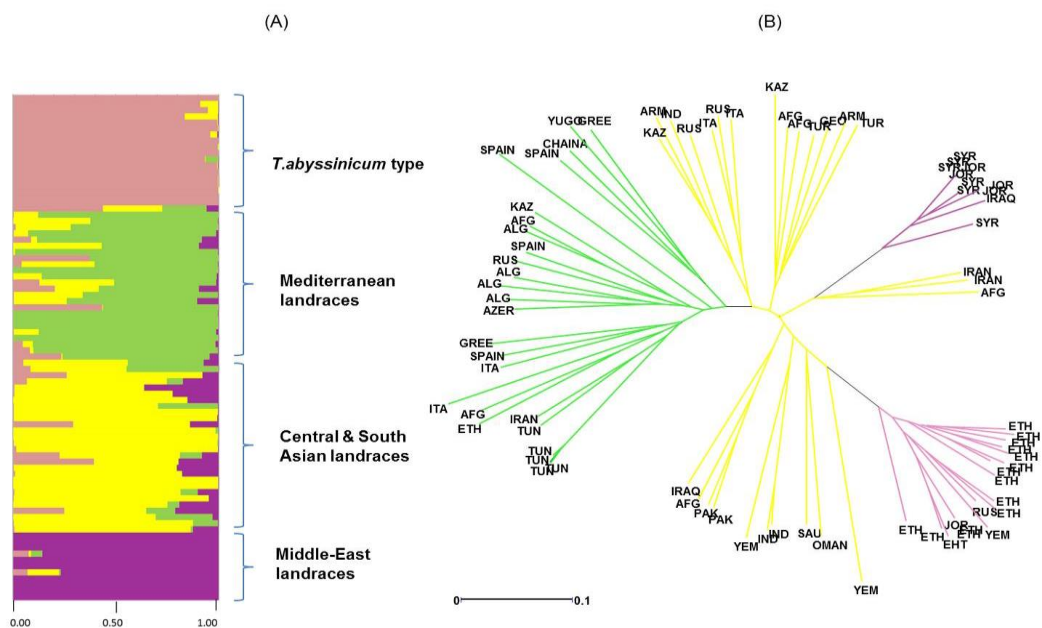

Figure II-5: Diversity in admixture among landraces by ad hoc STRUCTURE analysis. (A) Admixture analysis by kinship color coded following the colors of Figure 3 (B) Phylogenetic tree of evolutionary distances based on kinship values color coded following the colors set in figure 3

In fact, cluster 4 containing mostly Central and South Asian landraces had the highest level of admixture, it located toward the origin of the tree, and created one independent branch with seven landraces from Armenia, Kazakhstan, India, Russia, Afghanistan, Turkey and Georgia. Still, landraces from the same cluster also distributed to the branches generated by cluster 2 and 3. Cluster 2 containing mostly landraces from Ethiopia generated the branch further away from the origin of the tree, with landraces of cluster 4 from Iraq, Afghanistan, India, Pakistan, Yemen, Saudi Arabia, and Oman that located along this branch (Figure II-5). Cluster 1 generated an independent phylogenetic path, with landraces from Syria, Jordan, and Iraq composing the edge of the branch. Cluster 3 included landraces from many countries. The coordinates of the collection sites of the landraces were placed on a map and color coded to match their cluster assignment (Figure II-6). The admixture level among elites was higher than for the landraces (Figure II-4). The cluster containing ICARDA breeding lines was the largest group, and it is therefore unsurprising that it also presents the highest level of admixture, followed by the breeding program exchange subgroup. The cluster of Italian breeding program has noticeable allelic similarity with the cluster of Mediterranean trades, but also with the cluster of breeding programs from developed countries. Obviously, the cluster of Om Rabi derivatives is the smallest group, which shows a low level of admixture, but still it has some alleles in common with the cluster of ICARDA breeding program. Also the CIMMYT's elites group shared an important amount of alleles with the latest cluster.

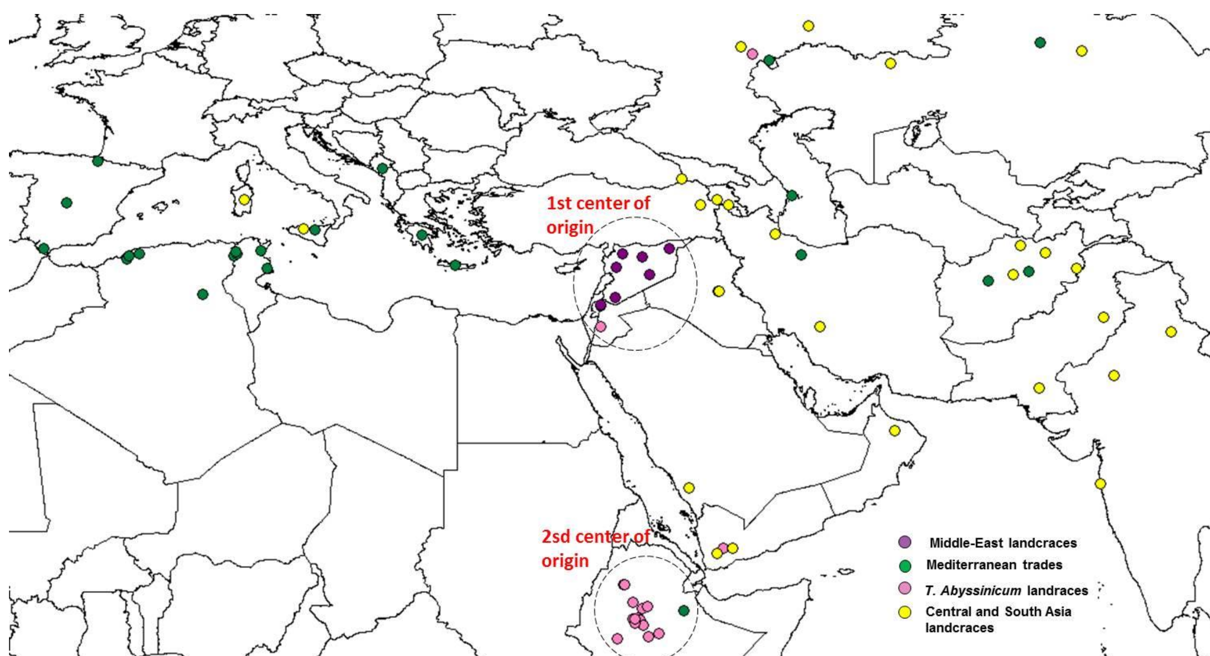

Figure II- 6: Geographical distribution of the coordinate of collection of the durum wheat landraces. Color codes are provided following the colors set in Figure 3 Dashed circles indicate the two centers of origin or diversity.

II-3-4. Genetic diversity among 10 sub-populations

A second set of 500 markers capturing the rarest alleles (0.005≤PIC≥0.01) and the full set of polymorphic markers were used to assess genetic diversity. Genetic diversity among clusters was investigating using these two sets of markers (Table II-4). Cluster 1 presented the lowest genetic diversity with PIC= 0.04, followed by Cluster 5 (PIC=0.05), these two clusters also had the lowest frequency of segregating markers 0.13 and 0.19, respectively, and did not capture any of the rare alleles. The group of Ethiopian landraces (cluster 2) had also a low level of segregating alleles (42%) comparable to what observed for breeding programs from developed countries (cluster 8), and equally captured very few rare alleles (1%). The Central and South Asia landraces was the most genetically diverse sub-population, with a PIC of 0.22, almost all markers segregating (0.84) and also captured 62% of the rare alleles. This was also the cluster with the highest admixture that distributed along all other phylogenetic branches. Among the modern germplasm, clusters 7 and 9 were the most diverse, with 58% and 51% of total alleles captured, and 4% and 3% of rare alleles represented in the sub-populations, respectively. Interestingly, cluster 10 maintained the highest level of rare alleles (21%) among the modern germplasm.

Table II- 4: Genetic diversity per sub-population.

Cluster ID	N	All markers			Rare alleles		
Name		Fixed[a] (%)	Segregating (%)	PIC[b]	Fixed[a] (%)	Segregating (%)	PIC
1. Middle East	11	0.77	0.13	0.04	1.00	0.00	0.005
2. *T. abyssinicum* type	18	0.47	0.42	0.13	1.00	0.01	0.005
3. Mediterranean trades	26	0.35	0.65	0.17	0.94	0.06	0.005
4. Central and South Asian	27	0.15	0.84	0.22	0.38	0.62	0.005
5. ‘Om Rabi’ derivatives	13	0.78	0.19	0.05	1.00	0.00	0.005
6. Italian	26	0.53	0.46	0.11	0.99	0.01	0.005
7. Breeding programs exchange	58	0.42	0.58	0.14	0.96	0.04	0.005
8. Developed countries	30	0.56	0.43	0.12	0.99	0.01	0.005
9. ICARDA	119	0.48	0.51	0.11	0.97	0.03	0.005
10. CIMMYT	42	0.56	0.44	0.10	0.79	0.21	0.005

[a] Refers to major allele

[b] Polymorphism information content

II-4. Discussion

Genetic diversity is of paramount importance as a source of novel traits and alleles for plant breeding, particularly to face the unpredictable challenges laying ahead, at a time of changing climates and new end-user demands (Tester et al., 2010). However, diversity *per se* is of limited use (Frankel et al., 1995: Royo et al., 2009). It is instead to the breeders' advantage to know which ideal sources of diversity should be integrated within each program. With this scope, the global diversity of durum wheat was assessed comparing breeding efforts, historical cultivars, and landraces from 28 countries. Genotyping with the Axiom 35K "breeders' array" revealed that the panel used could capture 36% of total polymorphism existing for the A and B genomes markers. A subset of 500 highly informative SNP markers was used to assess the genetic structure and stratification of the panel. This number of markers was in excess when compared to what previously reported for wheat with 70 simple sequence repeats (SSR) by Maccaferri et al. (2003), 97 SSR by Royo et al. (2010), 178 diversity array technology (DArTs) markers by Raman et al. (2010), and 469 SNPs by Cabrera et al. (2015). Overall, the genotyping results were satisfactory and allowed the implementation of all downstream applications. Likewise, the core subset panel used was able to capture a significant amount of the total diversity of durum wheat.

II-4-1. Success level of the clustering procedure

Human practices such as farming, consumption habits, and trading of seeds within and among communities generate pressures, drift or founder effects on the germplasm (Dyer et al., 2008; Deletre et al., 2011). Furthermore, societal, cultural, and natural barriers reinforce reproductive isolation, limiting or encouraging gene flow among cultivars (Deu et al., 2008; Pusade et al., 2009). Thus, several factors can influence the genetic diversity within a germplasm collection and the analysis presented here can only explain a fraction of it. The results of the AMOVA confirmed that the DAPC model was able to capture approximately one fifth of the total variance by stratifying the panel into 10 clusters, with individuals that maintained high levels of genetic diversity within groups. Thus, even if the choice of k = 10 was conservative as shown by the AMOVA, it was considered adequate to better identify similarities between genotypes, rather than over-fit their differences (Jombart et al., 2010). Clustering landraces by allelic similarities (kinship and admixture) is *de facto* an attempt of tracing those alleles that are identical by decent, maintained from their original domestication event or shared environmental pressures. In fact, landraces rarely hybridize with each

other because of the geographical distance and the self-pollinating nature of wheat, thus tend to share very limited allelic similarities. The exception is therefore represented by those alleles that are maintained since the common initial origin of the germplasm. Instead, in the case of cultivars and elite lines, genetic similarity is strongly influenced by the breeders' subjective choice of hybridizing specific germplasm sources to develop new lines. Since different breeding programs tend to utilize the same founders in their crossing strategies, strong admixtures exist between geographically distant germplasm. Separating cultivars and elite lines into groups of shared allelic similarity is therefore an attempt to capture the complex hybridization history of the breeding germplasm.

The population stratification of this panel suggests four sub-populations of landraces and six of modern germplasm. DAPC and STRUCTURE agreed on the number of clusters (k), but 4.3% of genotypes were not assigned to the same clusters by two or more of the software used for the analysis (Table II-3 and annex 1). In contrary neighbor-joining analysis suggested only 6 main clusters (Figure II-7). Instead, STRUCTURE analysis needed extra fitting to subdivide the landraces into separate clusters. Considering the high level of admixture showed by the landrace sub-group identified by STRUCTURE, and the clustering by UPGMA tree (Figure II-7), the results of DAPC were deemed more reliable. In fact, once landraces and elites were analyzed separately it was possible to reach good agreement between DAPC and STRUCTURE.

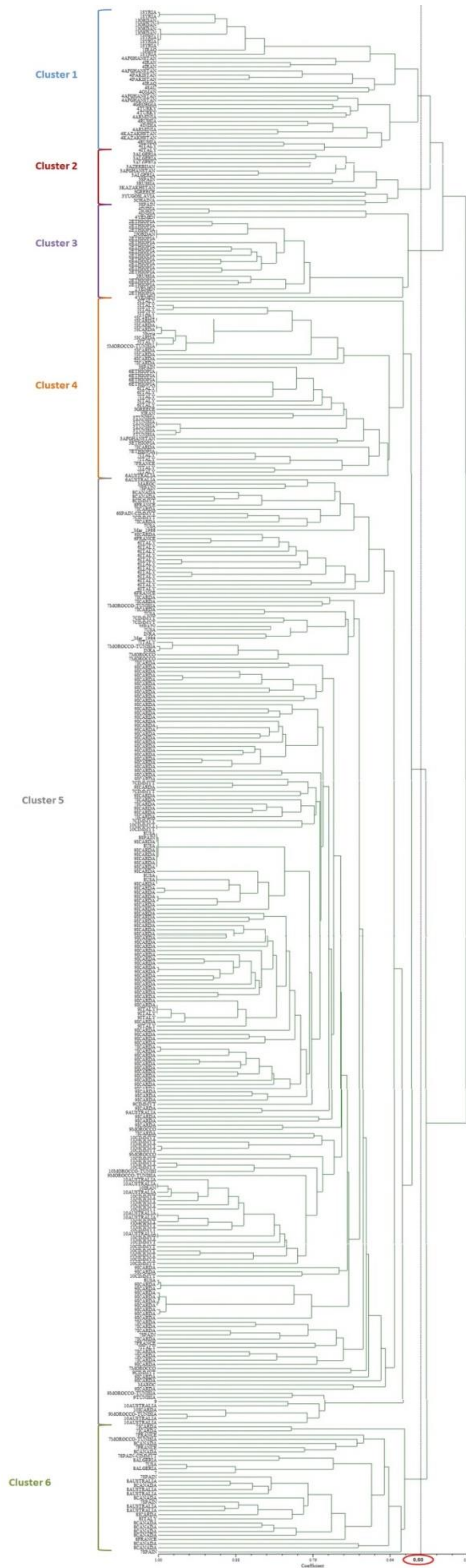

Figure II- 7: Neighbor-joining tree based on SNPs data using Rogers' genetic distance for 370 lines of durum wheat. Genotypes are coded based on the output of DAPC clustering. And clusters are defined using the vertical line in black as reference.

II-4-2. Clusters of genetic similarity

As described above, the population stratification of this panel separates four groups of landraces and six groups of cultivars and elite lines. Cluster 1 includes landraces from Jordan, Syria and Iraq (Figure II-5), countries that correspond to the center of origin of durum wheat in the Levant (Mackey, 2005). The geographical proximity of these landraces to the center of origin maintained a high level of genetic purity with low levels of admixture (Figure II-5) and almost completes fixing of major (77%) and minor (10%) alleles (Table II-4). This is in good agreement with what was reported previously for landraces from Jordan (Rawashdeh et al., 2007; Mohammadi et al., 2014). More interestingly, the phylogenetic tree (Figure II-4) clearly indicates how the germplasm from Syria and Jordan are more closely related as compared to the germplasm from Iran and Afghanistan belonging to cluster 4. This would suggest that durum wheat truly originated in the South end of the Fertile Crescent (Mackey, 2005), and only later migrated to the neighboring regions. Cluster 2 groups mostly landraces from Ethiopia, with the exception of one from Yemen, one from Jordan and another from Russia. However, the Russian landrace appears as wrongly assigned based on its high level of admixture (Figure II-5). Ethiopia is known as a "secondary center of durum wheat diversity" (Harlan, 1969; Vavilov, 1992). Landraces from this country have unique morphology (Sakamoto and Fukui, 1972; Porceddu et al., 1973; Pecetti et al., 1992) and represent a separate sub-species under the name *T. durum* subs. *abyssinicum* or *T. aethiopicum* (Dejene et al., 2015; 2016). Figure 5 clearly shows that this germplasm is distinct from the primary region of origin of durum wheat (Middle-East landraces) with substantially no kinship to it. Furthermore, there is limited admixture between this group and any others. Hence, Ethiopia truly represents a different center of diversity for durum wheat, without an evident allelic similarity to the primary origin in the Levantine, as also suggested by several other authors (Sakamoto and Fukui, 1972; Porceddu et al., 1973; Pecetti et al., 1992). The lack of allelic similarity between the two centers of diversity can be due to adoption in Ethiopia of a population of landraces from the Middle East that was genetically different from those that can be found there today (founder-migration effect), or as a separate domestication of *T. dicoccum to T. durum*. While both explanations are valid, the tight geographical distribution, low admixture observed among the landraces from the Levant do not support the hypothesis that a population of landrace existed within this region and then migrated to Ethiopia. However, caution should be used here as one Jordanian landrace was

identified among this cluster with 99% of genetic similarity to *T. abyssinicum* types. This landrace does not show the traditional relaxed spike morphology of the *T. abyssinicum* type (Figure II-8) so it remains extremely hard to conclude if this indeed represents the one original landrace population that migrate from the Levantine to Ethiopia. Certainly, its complete lack of genetic similarity to the other Levantine landraces seems to suggest that it was more likely migrated from Ethiopia back to the Levantine, rather than the opposite direction. The second hypothesis then appears slightly more plausible. Domesticated emmer reached Ethiopia some 5,000 years ago (National Research Council, 1996) probably arriving from Egypt along the Silk Road (Luo et al., 2007) and it occupies approximately 7% of the wheat production today under the name of *aja*. Thus, it can be suggested that Ethiopia is indeed not a secondary center of diversity, but rather a "secondary center of origin", where emmer was further domesticated to durum wheat as it occurred in the Levantine more than 7,000 years before.

Cluster 3 unifies the landraces from the Mediterranean basin (Italy, Greece, Tunisia, Algeria and Spain), plus few originating from Ethiopia, Afghanistan, Azarbijan, Kazakhstan, China, Yugoslavia, Iran, and Russia. Foremost, the landrace collected from Ethiopia that belong to this group do not show the typical morphology of *T. abyssinicum* (data not shown). Therefore, it is possible that even if these were collected in Ethiopia, they might not be of the *T. abyssinicum* type. A simple explanation might be that these are rather historical Italian cultivars derived from landraces (Scarascia, 2005) and brought to Ethiopia during the occupation by Italy from 1936 to 1941. Concerning the other countries, it is possible that these landraces are also of Italian origin and from there spread to the neighboring countries through trading in the last millennium. However, with the exclusion of the similarity between Italian and Ethiopian landraces, in the phylogeny tree all genotypes branch out directly from the root of the cluster (Figure II-5B), which indicates the existence of strong genetic differences among individuals. In fact, this population holds high level of genetic diversity (PIC=0.17), low levels of alleles fixation (35%) and high level of admixture. Thus, the hypothesis of one single common origin from Italy appears unlikely. In fact, their shared allelic identity suggests that these might have originated from related seed sources, which have then be exposed to similar natural pressures by the environment.

Cluster 4 is the largest among landraces, the most genetically diverse overall, and it clusters entries from 18 countries (Annex 1). It also shows severe levels of admixtures

between the three clusters described above (Figure II-5). These landraces are therefore likely the result of migration and hybridization of germplasm belonging to clusters 1, 2 and 3. This cluster can be further divided into 4 sub-populations based on their admixture levels (Figure II-5). The sub-groups 1 and 2 are derived primarily from Central Asia (Kazakhstan, Afghanistan, India, Armenia, Turkey, Russia, Georgia and Italy). They are evolutionarily closer to the “Mediterranean types”, but maintain good distinction with little similarity among individuals. As described for Cluster 3, this type of similarity with strong individual diversity is probably best explained as multiple separate sampling events from a common seed source, combined with shared environmental pressures. Thus, a scenario can be devised where merchants or tribes departing the Levantine for Central Asia carried with them seeds from similar field sources. The third sub-group includes entries collected in the Fertile Crescent (Iraq, Afghanistan and Pakistan), South Asia (Yemen and India), and the Arabian Peninsula (Oman, Saudi Arabia, and Yemen). This group locates evolutionary along the branch of *T. abyssinicum* types. In particular, the landraces from Yemen and Oman are more closely related to the Ethiopian landraces, and are therefore the probable result of dispersion from this secondary center of origin or domestication of durum wheat. The last sub-group is composed of landraces from Iran and Afghanistan located in the same branch with Middle-East landraces, and thus likely dispersed from here.

A special note is required for the landraces of Russian origin. These were identified in clusters 2, 3, and two sub-groups of cluster 4. This level of genetic diversity is normally unexpected for a country so geographically distant from the two centers of origin of durum wheat. However, this vast region has witnessed large migration since its origin, a well-documented source of genetic variation (Vavilov, 1951; 1992). Eleven landraces from the core subset were grouped within the clusters of cultivars and elite lines. The simplest explanation is that these were not true landraces, but rather old tall cultivars that were wrongly labeled during the collecting missions by the gene banks. Alternatively, landraces have often been considered a key resource for contemporary agriculture and thus have been used in plant breeding programs to enlarge the genetic diversity of modern genetic pools (Bradshaw et al., 2005; Sharma et al., 2013). Hence, it is possible that these landraces are among those utilized in recent years to improve biotic and abiotic tolerance, or were favored by the early breeders like Nazareno Strampelli to develop pure lines (Scarascia, 2005).

Cluster 5 is small, composed of just 13 elite lines and cultivars from ICARDA breeding program and most of them include the cultivar ‘Om Rabi’ in their pedigree. Om Rabi is the name of the largest river of Morocco, and this name was attributed to one of the first cross ever produced by the ICARDA durum breeding program in 1981, which combined the widely cultivated Jordanian landrace ‘Haurani’ with the successful CIMMYT line ‘Jori69’. Cultivars have been released in 12 countries from this cross under various names (‘Cham 5’, ‘Tomouh’, ‘Om Rabi’, ‘Oum Rabi’, ‘Omrabi’, ‘Gahar’, ‘Um Qais’, and ‘Aydin93’) and they remain widely cultivated by smallholder farmers in the driest areas of central West Asia and North Africa. Considering that 50% of the genome of this cross is derived from a landrace, it is not surprising that it shares admixture with Cluster 1, and it has very similar allelic fixation as the landraces from the center of origin of durum wheat (Table II-4).

Cluster 6 brings together modern and old cultivars developed by Italian breeders. Substantially this set of lines is derived from the initial work of Nazareno Strampelli and the following “fathers” of Italian breeding (Scarascia, 2005). The admixture level is high (Figure II-4) and it captures 46% of the total alleles assesed (Table II-4), indicating that several breeding programs worldwide utilized the work carried on in Italy as a base of their cultivar development pipeline. However, the level of genetic diversity is low compared to other breeding clusters (PIC=0.11), which could be the result of the frequent use in hybridization of a reduced number of founders, in combination with strong selection pressure for the same traits needed for the Italian growing conditions and the rheological requirements of the pasta industry.

Cluster 7 is located at the center of the graph of the two main IPC by DAPC (Figure II-3) and it groups together material from several countries and breeding programs such as Spain, Morocco, ICARDA and CIMMYT. It originates from the sharing of several germplasm sources among breeders targeting dryland agriculture. Thus, the genetic similarity between this germplasm can be explained as the common origin of allelic sources, together with the imposition of similar selection pressure for specific dryland-traits. This cluster has the highest rate of genetic diversity (PIC = 0.14), portion of captured alleles (58%), and frequency of rare alleles (4%) than any other cluster of modern germplasm. In addition, the high admixture (Figure II-3) and central position in the DAPC graph (Figure II-3) confirm that this cluster is the founding base that guarantees good exchange of alleles among all other breeding programs.

Germplasm from USA, Australia, and Canada were grouped together in cluster 8, together with four lines from Italy, Spain and France and two landraces from Algeria. This cluster captures the least amount of available allelic variation (43%) or rare alleles (1%) among breeding programs, and one of the lowest PIC (0.12). Considering the geographical distance between the breeding programs grouped here, and the different environmental conditions, it is a good example of the decay in genetic diversity that other authors have suggested is occurring in breeding programs worldwide (Hoisington et al., 1999; Martos et al., 2005). The tight rheological requirements imposed by the pasta industry has pushed durum wheat breeders to maintain their hybridization programs extremely narrow, using often the same set of standard cultivars as donors of quality traits (Karagöz et al., 2005; Zencirci et al., 2005; Altintas et al., 2008). This is reflected by the high number of fixed alleles identified in this cluster (57%) for which genetic diversity no longer exist within these breeding programs. Still, it is important to also indicate that for a large portion of the genome (43%) genetic diversity was captured and can be exploited to make further genetic gain.

Cluster 9 groups together the vast majority of the germplasm of ICARDA included in this analysis, with the exclusion of the 'Om Rabi' derivatives assigned to Cluster 5, and some of the genotypes included in Cluster 7. The durum wheat breeding program of ICARDA officially started in 1981 and run for over 20 years under the umbrella of CIMMYT. This program primarily targeted drylands agriculture using crossing schemes involving both modern and primitive germplasm. It released over the years 99 cultivars in 24 countries (Latican et al., 2015). Within this group are also included some of the Italian cultivars derived from 'Creso', a radiation mutant of Strampelli's cultivar 'Capelli', and several of the CIMMYT lines derived from 'Yavaros', a CIMMYT cultivar that spread widely in North Africa and is today the most grown in Morocco, Algeria, and Tunisia under the name of 'Karim' (syn. 'Bittern'). The genetic similarity between 'Creso', ICARDA's and CIMMYT's materials can be found in the pedigree of ICARDA's breeding lines, which widely used 'BiCre' as a parent. In fact, 'BiCre' is derived by the simple cross of 'Bittern' and 'Creso'. This cluster captures 51% of the total allelic diversity available in the panel, and 3% of the rare alleles. Due to the large size (N=119) the PIC is low. This shows that even a breeding program that specifically targets genetic diversity as an adaptation strategy via frequent crosses to primitive germplasm can erode large parts of it by exposing the germplasm to severe selection in challenging environments. Still, this cluster is the second most genetically wide among

modern germplasm. Also, ICARDA's breeding lines spread over other two clusters, thus meaning that overall the program maintained acceptable levels of diversity.
The breeding program of CIMMYT has been running for over 50 years. It has had the ability to deliver superior cultivars throughout the developing world and still serves today as source of useful alleles for industrialized countries. As for the ICARDA's cluster, the severe selection pressure during breeding caused a shrinkage in the overall genetic diversity, with the vast majority of the CIMMYT's germplasm clustering mostly in one group (Cluster 10) with the lowest PIC (0.10) and 56% of the genome in fixed status. However, this cluster also captures 21% of the available rare alleles, which is by far the best achievement in that sense among breeding programs. Furthermore, CIMMYT breeding lines can also be found in Clusters 9 and 7, which suggest an overall high level of allelic diversity remains available for breeding advancements.

Figure II- 8: Spike morphology of the two centers of origin's landraces. (A) Middle East landraces, (B) *T. abyssinicum* type.

II-4-3. Comparison to other population stratification studies of durum wheat

In previous research, a panel of 190 Spanish durum wheat landraces was attributed to nine sub-populations (Ruiz et al., 2012), while a similar set of the germplasm collection used here, comprising 134 modern durum cultivars, was assigned to six sub-populations by Maccaferri et al. (2003). In our research, the number of clusters used for stratification could have been increased, to allocate four additional sub-populations among landraces of Cluster 4. However, the setting of k is highly dependent on the scope of the research and here the preference was given to capturing similarities rather than divergences. A large portion of admixture among landraces remained unfixed with the set value of *k*, and this could justify the difference in the number of clusters between our work and that of Ruiz et al. (2012). The ability to distinguish following waves of dispersion among landraces was one of our scopes and this was achieved by finding separation between the two main centers of origin/diversity (Middle East and Ethiopia) and other landraces. Similarly, the division into six clusters of modern material appeared in line with the results of previous authors (Maccaferri et al., 2003) and it provided interesting information about the history of alleles exchange among breeding programs. Slight differences were, however, observed from past works, due to the significant increase in this study of the number of elites derived from the ICARDA breeding program, which alone defined two novel well distinct clusters (5 and 9), and also the study of recent Australian and Canadian cultivars, which also created a cluster not described before by Maccaferri et al. (2003).

II-4-4. Genetic diversity and the future of durum wheat breeding

The scenario of today's global cultivation of durum wheat can be summarized in the work of few great breeders: North Africa and the Middle East countries still heavily rely on 'Karim' (syn. 'Bittern', 'Yavaros79') a mega-cultivar bred by Dr. Gregorio Vazquez (CIMMYT), similarly 'Simeto' is probably the most grown cultivar around the World and it was developed in Italy by Dr. Fortunato Calcagno (Pro.Se.Me), and the 'Cham' series that occupy some of the driest areas of the World were bred at ICARDA in Syria by Dr. Miloudi Nachit. The scenario of the industrialized World is only slightly more segmented, with few mega-cultivars also occupying significant land area. As shown by this genetic diversity study, most of the modern germplasm has only 58 to 44 % of the genes still segregating, regardless of the breeding strategy or combination of germplasm utilized. Unexpectedly, the two centers of origin of durum wheat do not

appear to be the most exploitable source of allelic diversity with most of their loci in fixed state. Rather, the landraces from Central and South Asia revealed the highest accumulation of rare and normal alleles and should therefore be kept in high consideration for increasing diversity of modern breeding programs. Alternatively, the five clusters of ICARDA, CIMMYT, developed countries, 'Om Rabi' derivatives, and Italian breeding showed limited admixture with each other (Figure II-4) and therefore their inter-hybridization is a possible source of genetic diversity. Still, this will be possible only if the exchange of seeds for breeding purposes is kept free and unobstructed.

Chapter III: Heat Tolerant Loci of Durum Wheat Defined via a North-South Gradient

Hafssa Kabbaj[1,2], Amadou T. Sall[1,2], Madiama Cisse[3], Habibou Gueye[4], Mulatu Geleta[5], Abdelkarim Filali-Maltouf[2], Bouchra Belkadi[2], Rodomiro Ortiz[5], Filippo M. Bassi[1]*

1 International Center for Agricultural Research in the Dry Areas, Rabat, Morocco,
2 Department of Plant Science, Mohammed V University, Rabat, Morocco,
3 Senegalese Institute for Agricultural Research (ISRA), Saint-Louis, Senegal
4 National Center for the Agricultural Research and development (CNRADA), Kaedi, Mauritania
5 Plant Breeding, Swedish University of Agricultural Sciences, Alnarp, Sweden

Abstract

The production and productivity of durum wheat in most parts of the world is constrain by drought and high temperature stresses. Heat tolerance is a complex trait under the influence of several genes. In this study, 216 durum wheat modern lines were field tested in the South of Morocco and along the Senegal River to create a north-south heat gradient of 10°C. the environments were clustered on mega-environments based on weather and phonological data, then for each yield component: grain yield, 1,000 kernel weight, grain number per spike and grain number per meter square, the heat stress susceptibility index (HSI) was calculated between clusters. A total of 7,652 SNPs polymorphic markers were identified among the lines and were used to define 157 marker-trait associations (MTAs), corresponding to 25 QTLs for HSI. seven QTLs were considered as stable along the heat gradient, the QTL Q.icd.DW.01 on chromosome 1A was identified as highly significant with an LOD of 6.5, as well as Q.icd.DW.08 on chromosome 2B (LOD=4.8) and Q.icd.DW.23 on 7B with an LOD of 4.0 that were identified as multi trait loci. The identification of these critical heat tolerance loci unlocks the potential of deploying molecular techniques to quickly pyramid them into climate change-ready varieties.

III-1. Introduction

Durum wheat (*Triticum turgidum L. var. durum*) is a tetraploid species of wheat that is widely cultivated today (Blanco et al., 1998; Shewry, 2009). It is used in different parts of the world for several food products, such as pasta, couscous, unleavened bread, bulgur, and mote, among others (Nachit, 1992). The production and productivity of durum wheat in most parts of the world is constrained by raising temperatures (Lobell et al., 2011; Aseng et al., 2015). In particular, for wheat, temperatures in excess of 30º C during the reproductive phase can severely reduce grain numbers and dry weight (Wollenweber et al., 2003; Dias et al., 2010; Sall et al., 2018). Environmental temperatures have increased since the beginning of the last century and they are predicted to further increase under the present treat of climate change, (Hansen et al., 2012). Various climate models indicated that the global mean temperature will increase by 1–4 °C by the end of the 21st century (Driedonks et al., 2016). Hence, new wheat cultivars better adapted for future climatic conditions will therefore be required (Semenov and Halford, 2009; Foulkes et al., 2011; Zheng et al., 2012; Semenov et al., 2014; Sall et al., 2018). Breeding for heat tolerance cultivars is still in its infancy stage and warrants more attention in the near future (Ortiz et al., 2008; Ashraf et al., 2010). Besides, selecting for heat tolerance via yield trials in sites where the stress occurs is complicated by the complex interaction between genotype and environment (Reynolds et al., 2012). Though, traits like thousand grain weight (TGW), grain number, and canopy temperature have been proposed as more reliable and inheritable proxy (Reynolds et al. 2001), their use achieves only partial success.

A more reliable method could be the identification of molecular markers associated with yield and its components under high temperatures. Such genomic regions have been previously mapped using quantitative trait locus (QTL) analysis in bi-parental population for grain filling duration, thousand grain weight, yield, canopy temperature depression, stay green, and senescence related traits (Paliwal et al., 2012; Vijayalakshmi et al., 2010; Kumari et al., 2013; Mason et al. 2010; Pinto et al. 2010, Bhusal et al., 2017). However, their deployment in breeding thus far has been limited.

Several association mapping studies have been conducted to dissect the genetic basis of grain yield in durum wheat (Mengistu et al., 2016; Kidane et al., 2017). Recently Sukumaran et al., 2018, identified 112 markers trait association to heat stress in subset of 208 line of durum wheat from CIMMYT gene bank accessions that were previously

evaluated and characterized for use in breeding for heat and drought tolerance. The panel used by Sukumaran et al., 2018 was phenotyped under yield potential (YP), drought stress (DT), and heat stress (HT) conditions, where combined analysis of three environments (YP+DT+HT), and its comparison with trait per se and stress indices identified QTL hotspots on chromosomes 2A and 2B for yield components.
The aim of this study was first to define heat susceptibility indexes (HSIs) for several yield-related traits recorded on a panel of modern durum cultivars grown in the South of Morocco and under severe heat along the Senegal River. Second, to identify the loci responsible for guiding heat tolerance to be then deployed in breeding.

III-2. Materials and methods

III-2-1. Plant material

A subset of 216 durum wheat elite lines was selected out of a large panel of 384 lines as described in Kabbaj et al. (2017). The panel was genotyped with 35K Affymetrix Axiom wheat breeders array at Trait Genetics (Gatersleben, Germany) as described in Kabbaj et al. (2017) to derive 7,652 good quality polymorphic markers, with a minor allele frequency superior to 3%. The entries were assigned to 10 sub-populations population using the discriminant analysis of principal components (DAPC) algorithm implemented on the 'adegenet' package 1.4-1 (Jombart et al., 2010) in R Studio V 2.3.2 (R Development Core Team, 2011), while the kinship matrix was calculated via STRUCTURE v 2.3.4 (Pritchard et al., 2000). Full, details are presented in Kabbaj et al. (2017). The analysis identified six sub-populations: "Om Rabi derivatives" cluster, "Italian breeding program" cluster, "Breeding program exchange" cluster, "developed countries" cluster, "ICARDA" cluster and the cluster of "CIMMYT". Linkage disequilibrium (LD) decay was estimated by first aligning the marker sequences to the recently released bread wheat genome assembly (IWGSC, 2018), and then determining the unlinked distance by TASSEL V 5.0 software (Bradbury et al. 2007) at P-values ≤ 0.01 to define an LD of = 51.28 Mb (Sall et al., 2019 unpublished).

Out of 35k, 8,173 SNPs markers (36%) were found to be high quality and polymorphic. 7,542 were aligned to svevo (Maccaferri et al., 2019) with 98 to 100% of identity. The description of genotypic data and population structure are presented in the work of kabbaj et al., 2017.
Understanding the LD pattern in germplasm is important for selecting the marker density required for GWAS and for defining identified QTL regions (Siol et al., 2017).

The LD decay was calculated using Neanderthal method as discussed by Sall et al., 2018 where the LD values was 51.28 Mbp

III-2-2. Field experiment and phenotyping

Two irrigated field stations were used for phenotyping in the South of Morocco Tessaout (29.8348302, -8.5786262) and Melk Zhar (30.019832, -9.486629). In addition, two irrigated field stations located along the Senegal River were utilized to impose a strong heat stress as described in Sall et al. (2018) in Mauritania (Kaedi, 5.999592, -13.071842) and in Senegal (Fanaye, 16.538348, -15.221948). The genotypes were grown during seasons 2014-15 and 2015-16 at all sites, with the exception of Tessaout that was utilized only in 2015-16. The maximum temperature at flowering in Tassaout is 26.6 °C, during the two crop season in Melk zhar and fanaye was around 30 °C and 37°C, respectively. However, in Kaedi it was 39 °C during 2014/2015 and it raised up to 41°C during the following crop season.

The layout was an augmented design with four checks replicated 11 times, and sub-blocks of size 24. The entries were sown in plots 4.5 m^2 at a sowing density of 120 Kg ha^{-1}. Planting was started one week earlier in Melk Zehr and Tassaout compared to the stations in Sub-Saharan Africa. In Kaedi (Mauritania) a total of eight gravity irrigations were performed for a total of approximately 390 mm of water during the season. In Fanaye nine gravity irrigations were performed for an estimated 410 mm of irrigation water. In Melk Zehr drip irrigation was used to deliver approximately 420 mm of water. In Tessaout six gravity irrigations were performed for an estimated total of 420 mm of water. Pre-planting fertilization was homogenized across sites with 50 units of N, P, K each. Nitrogen was further applied in 2 splits of 50 units each at 4th leaf stage (Zadok scale of 2) and during stem elongation (Zadok scale of 3), with the exception of Melk Zehr where 10 splits of 10 unites each were used via fertigation.

The following traits were recorded at all sites as described in Sall et al. (2018) and Zaim et al. (2017): days to heading (DTH), grain yield (GY), 1,000 kernel weight (TKW), number of grain per spike (Gr.spk) and number of grain per meter square ($Gr.m^{-2}$). The number of grain per spike and number of grain per meter square were recorded only in 2016 season.

III-2-3. Statistical analysis

Each location in each year were defined as environments for a total of seven. Analyses of variance were performed and best linear unbiased estimator (BLUEs) were obtained

for each environment individually using 'lme4' package in R Studio V 2.3.2 (R Development Core Team, 2011). Combined analysis across environments were conducted by best linear unbiased estimator (BLUE) considering all sources of variations as fixed effects. Heritability for each trait was calculated according to the formula:

$$H^2 = 1 - \frac{MS(GxE)}{MS(G)}$$

Individual environments were clustered by principal component analysis (PCA) in Genstat software (Payne, 2009) on the basis of the temperature data, growing degree days (GDD), days to heading (DTH), and day length at sowing and flowering (DL) (Table III-2). Thereafter, combined BLUEs were estimated for each cluster of environments.

Genotypes tolerant to heat stress were identified implementing the heat susceptibility index (Fischer and Maurer, 1978) using the equation:

HSI = [1 − (Ys)/(Yp)]/[1 − (.Ys)/(.Yp)]

Where *Ys* and *Yp* are values of the wheat lines evaluated under stress and non-stressed conditions, respectively, and. Ys and. Yp are the mean values of all lines for the specific trait obtained under stress and non-stress conditions, respectively. The BLUEs calculated for the IPC clusters combining Tassaout and Melk Zhar were considered as the non-stressed, while the clusters of Kaedi and Fanaye were considered as heat stressed. HSI was determined for GY, TKW, and Gr.spk.

Marker-trait-association (MTA) analyses were carried out using the TASSEL 5 software (Bradbury et al. 2007), where mixed linear models (MLM) were fitted using 7,652 SNP markers, population structure matrix (Q), and kinship matrix (K). The Bonferroni correction for 230 genome-trait hypothesis (as calculated by dividing the durum wheat genome of 12,000 Mb by the LD of 51.2 Mb) was determined at LOD of 3.3 for $p<0.01$. Individual LOD for each MTA was calculated by TASSEL based on the Q-Q plots and distribution of p-values (Sukumaran et al., 2012; Sukumaran and Yu, 2014). First, GWAS was conducted for GY, TKW, Gr.spk and Gr.m^{-}2 per cluster using DTH as co-variate. Secondly, the analysis was conducted for the HSI of the traits mentioned above, and only MTA that appeared in both analyses at a LOD above the Bonferroni cut-off are presented here. Finally, MTAs were grouped into individual QTLs if falling within distances inferior to twice the LD (104 Mb) from each other.

The conversion of Axiom markers to Kompetitive allele specific polymerase chain reaction (KASP) was achieved by LGC Genomics, which is is a fluorescence-based genotyping assay of polymerase chain reaction. This method enables bi-allelic scoring of SNPs at specific loci. The array sequences of the markers associated to desirable traits (MTA) were submitted to LGC Genomics for in-silico design of KASP primers using their proprietary software. Those that passed the in-silico criteria were purchased and used to genotype the independent validation set. For each marker that amplified and showed polymorphism, the regression cut off between phenotype and haplotype was imposed at r = 0.105 following Pearson's critical value (Pearson, 1985). Each KASP marker was tested for association with target trait. The top 20 and worst 20 lines were considered as the true positive and true negative for heat tolerance in the case of heat study and for drought tolerance in the case of root study. Hence, the accuracy was calculated as the ratio of the correct allelic call among all, sensitivity as the ratio of the correct positive allelic among the top 20 yielding lines, and specificity as the ratio of the correct negative allelic calls among the 20 worst yielding lines.

III-3. Results

III-3-1. Yield components performance under north-south heat gradient

For yield components traits, the crop season 2015 was better than 2016, this is due the field management and the raise in temperature during the second crop season either in Moroccan stations and Senegal River stations, (Figure III-1). The effect of water, temperature and soil type led to very highly significant differences ($P < 0.001$) in GY, TKW and Gr.spk across sites, both the genotype and genotype-by-environment ($G \times E$) effects were highly significant ($p < 0.001$) (Table III-1). Even though the environmental effect explained more than 99% of the total variation, it was still possible to identify genotypes that were significantly superior in grain yield, TKW, Gr.spk, Gr.m^2 across environments.

The top yielding station was Tessaout 2016 with an average yield of 12 t/ha and lowest yielding station was Fanaye 2016with an average yield of 2 t/ha. The maximum yield was observed in Melk zhar 2015 and riched up to 9 t/ha. While maximum variation was observed in Melk zhar station (5.5 and 6.5t/ha of difference between the top yielding and lowest yielding line in two crop season 2015 and 2016 respectively). Several lines were good and performed well across all sites for all the traits, such as IDON37-033, IDON37-036, IDON37-053, IDON37-062, IDON37-081, IDON37-141 and

DURUM_PANEL_UNIBO-024, but also many lines didn't show good performance under heat conditions for example: IDON37-071, IDON37-082, DURUM_PANEL_UNIBO-0172, DURUM_PANEL_UNIBO-0180, DURUM_PANEL_UNIBO-0181, MCHCB-0213 and FIGSDRYWET134.

some lines showed higher performances in terms of GY, Gr.spk and Gr.m-2 in north side compared with the south side, while when looking at TKW, it was higher under stressed conditions than normal conditions, we cite DWAyT-0215, DURUM_PANEL_UNIBO-013 and DURUM_PANEL_UNIBO-0264, at the main time

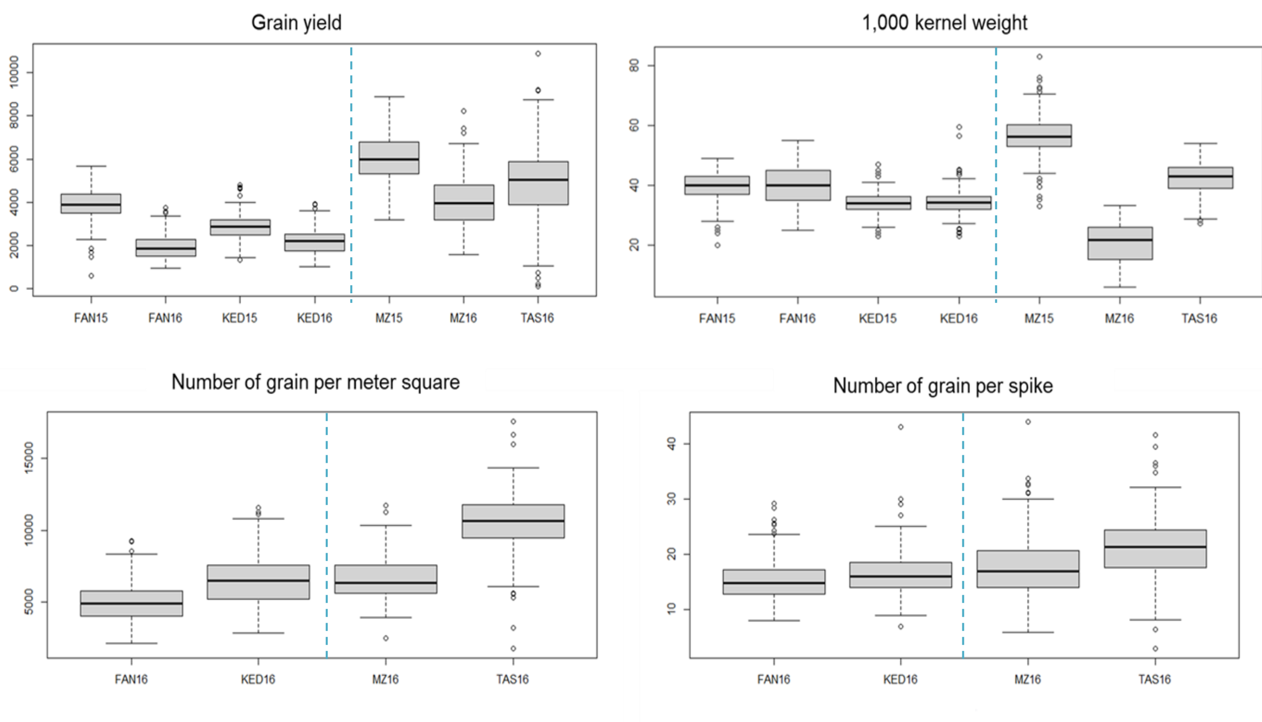

Figure III- 1: Boxplot showing the variation in grain yield (GY), 1,000 kernel weight (TKW), grain per spike (Gr.spk) and number of grains per meter square (Gr.m-2) for each location (Fanaye, Kaedi, Melk Zhar and Tassaout) during the two crop seasons 2014-15 and 201

Table III- 1: The analysis of variance of yield component trait (ANOVA)

TRAIT	Mean sq [G]	Mean sq [E]	Mean sq [GxE]	H_2
GY	1263893***	344471841***	673841***	0.41
TKW	76.8***	3658.8***	36.7***	0.52
Gr.spk	87***	39829***	55***	0.37

The heritability was in general good for all the traits. As expected, the site exposed to the highest temperatures (Kaedi) yielded the least. Instead Fanaye, which is less heat stressed, reached better grain yields, whereas the yield potential was 9.34 t ha^{-1} at Melk Zehr and up to 8.67 t ha^{1} in Tassaout, which had the coolest temperatures. The most surprising result was the top grain yield achieved along the Senegal River; i.e.; up to 5.7 t ha^{-1} in Fanaye or 4.8 t ha^{-1} in Kaedi with crop seasons of just 92 days in length.
The seven environments were clustered by principal component analysis (PCA) using the temperature information, average days to heading, average day length at the time of flowering, and growing degree days to flowering (Table III-2). Melz Zhar 2015 and 2016 were clustered together with Tassaout 2016, Kaedi 2015 and 2016 composed the second cluster, where the two crop season of Fanaye grouped in cluster 3 (Figure III-2). When comparing the Moroccan cluster against the two obtained for the Senegal River, GY and Gr.spk were higher, while TKW was lower.

Table III- 2 : Temperature data at flowering time (Maximum, minimum and average), growing degree days (GDD), day length (DL) at sowing and flowering times and days to heading (DTH) per location showing the difference between the two sides (Morocco and Senegal River)

Location	Max temp	Avg max	min temp	Avg min temp	GDD	DL sowing	DL flwr	DTH
Tessaout 2016	26.6	20.9	-0.2	4.9	1420.5	605.2	72702	110.6
Melkzhar 2016	30.4	22.3	2.0	7.5	1477.9	618.6	63856.1	98.0
Melkzhar 2015	31.4	20.4	0.4	6.7	1356.4	635.1	63322.5	100.1
Fanaye 2016	37.1	32.7	9.5	14.4	1593.8	669.0	46250.4	68.0
Fanaye 2015	37.9	31.1	8.3	15.0	1495.7	671.2	43870.8	65.4
Kaedi 2016	39.0	33.4	15.0	21.6	2019.6	671.6	50321.6	73.2
Kaedi 2015	41.0	31.0	10.0	18.0	1885.7	673.6	45973	67.4

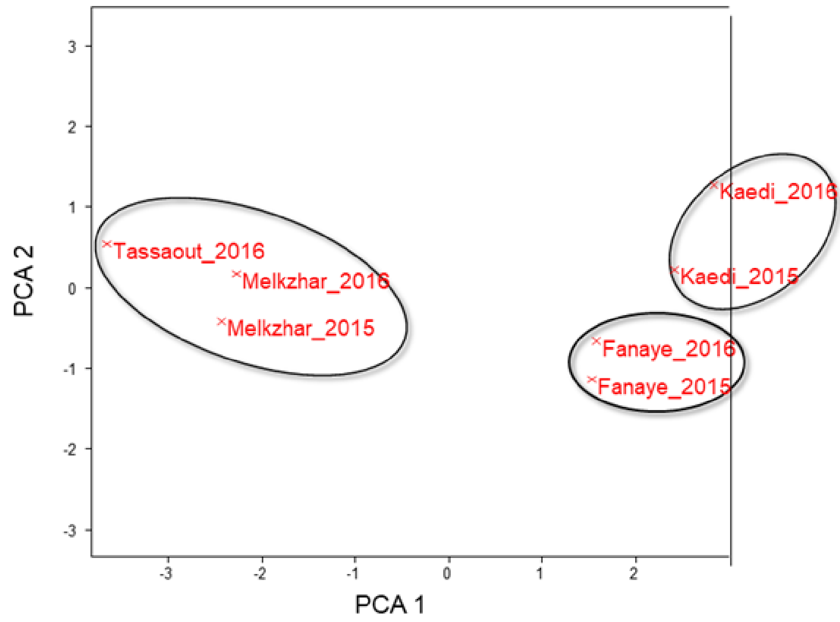

Figure III- 2 : Clustering of seven environments s by Principal component analysis (PCA) using the weather and phenological data presented in Table III-2. Black circles indicate clusters of similar environments.

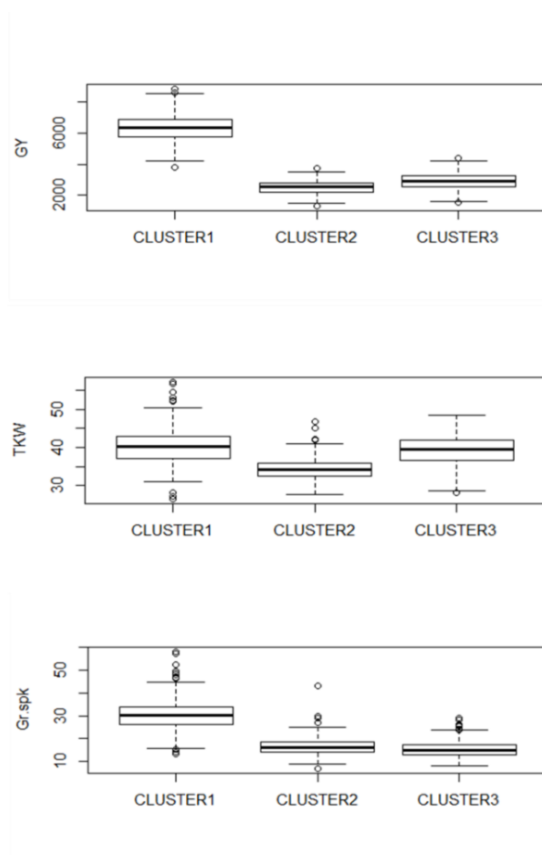

Figure III-3: Agro-morphological performances of the elite lines per clusters for GY, TKW and Gr.spk.

Table III- 3: Comparison of grain yield (GY), 1,000 kernel weight (TKW) and number of grains per spike (Gr.spk) between clusters defined in figure2

	GY			TKW			Gr.spk		
	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg
Cluster 1	8812.7	3824.2	6352.1	57.2	26.5	40.4	58.0	13.1	30.8
Cluster 2	3750.1	1347.0	2531.5	46.7	27.5	34.3	43.0	7.0	16.4
Cluster 3	4417.4	1541.6	2922.6	48.5	28.0	39.2	29.2	8.0	15.5

III-3-2. Heat susceptibility index (HSI)

The identification of heat tolerant and heat susceptible genotypes were conducted by comparing the two clusters of stressed environments (Kaedi and Fanaye) against the unstressed one (Tessaout and Melk Zhar), knowing that there was between 6 and 10°C temperature difference between them. According to Fisher and Maurer (1978), the entries with HSI < 0.5 are highly heat tolerant, HSI=0.51-0.75 are heat tolerant, HSI> 0.75 are considered as heat susceptible.

In Figure III-4, the number of tolerant and susceptible genotypes is presented. For GY we identified 18% of heat tolerant and 82% of non-tolerant lines. For TKW 28% of heat tolerant and 72 of heat susceptible lines, for Gr.m^{-2} 43% of tolerant lines versus 57% of susceptible ones, while for Gr.spk 45 and 55% were the percentage observed of tolerant and non-tolerant lines respectively. Overall, 13% of the lines were considered as highly tolerant for all four traits combined. In particular, IDON37-013, IDON37-081 and IDON37-097 were the top ones.

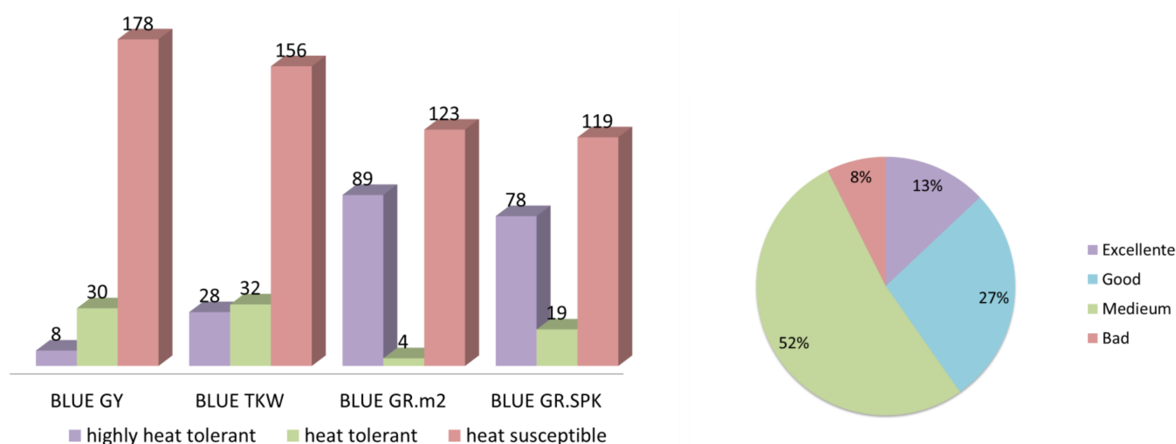

Figure III- 3: The combined heat susceptibility index (HSI) for GY, TKW, Gr.m-2 and Gr.spk showing the number of heat tolerant and heat susceptible lines of each trait in the left, and the total percentage of excellent, good, medium and bad lines for all traits.

III-3-3. Marker-trait associations (MTAs) for yield component and heat Susceptibility Index

A total of 63 MTAs were detected using the individual traits in each cluster and imposing DtH as covariate at -log P >3.3. The LOD (-log P) ranged from 3.3 to 9 and markers explained between 1% and 20% of the phenotypic variation.

The highest numbers of MTAs were detected for TKW (46) followed by Gr.spk (10), GY and Gr.m^2 (6, 1, respectively). Together the 63 MTAs landed on 25 genomic regions spaced less than twice the LD distance. Most of these QTLs were associated to multiple traits and were identified in two or more clusters.

Using the HSI values calculated between clusters, 157 MTAs were detected at -log P >3.3. Among these, 29 overlapped with genomic regions identified in the previous analysis and were then considered for downstream studies.

The highest numbers of MTAs were detected for HSI-GY (54) followed by HSI-TKW (43) and Gr.m2 and Gr.spk (33 and 21, respectively). The chromosomes 5B, 6B and 7B had the highest number of MTAs (22 each) followed by chromosome 2A with 16 MTAs, where chromosomes 4A has the lowest numbers of MTAs (1). The B genome had the highest number of MTAs (105) compared with A genome (52).

157 MTAs results in 25 QTLs associated with HSI for GY, TKW, Gr.m2 and Gr.spk overall the comparison made between Morocco and Senegal River.

Most of the QTLs found were consistent as they were overlapping the different combinations, of HSI made in 2015 and 2016, and were detected in the combined blues. In chromosome 1A, Q.icd.Heat.02 (LOD=6.46 and r^2= 18%) was identified in stressed environments, associated with GY, Gr.spk and their HIS. The QTL Q.icd.Heat.05 (LOD=4.11, r^2= 9%) in chromosome, associated as well with GY and Gr.spk and their HSI 2B but was detected in normal environment and stressed ones.

While in chromosome 5B the QTL Q.icd.Heat.30 (LOD=6.45, R2= 15%) was associated to GY, Gr.m^{-2}, TKW and the heat index , presented in stressed conditions only (Table III-4). Interestingly, 114 MTAs were found under stressed environments presented by 18 major QTLs (Circoss presentation figure III-5, and table III-4).

* Presence of the trait non-stressed environments
* Presence of the trait stressed environments
* Presence of the HSI of the trait

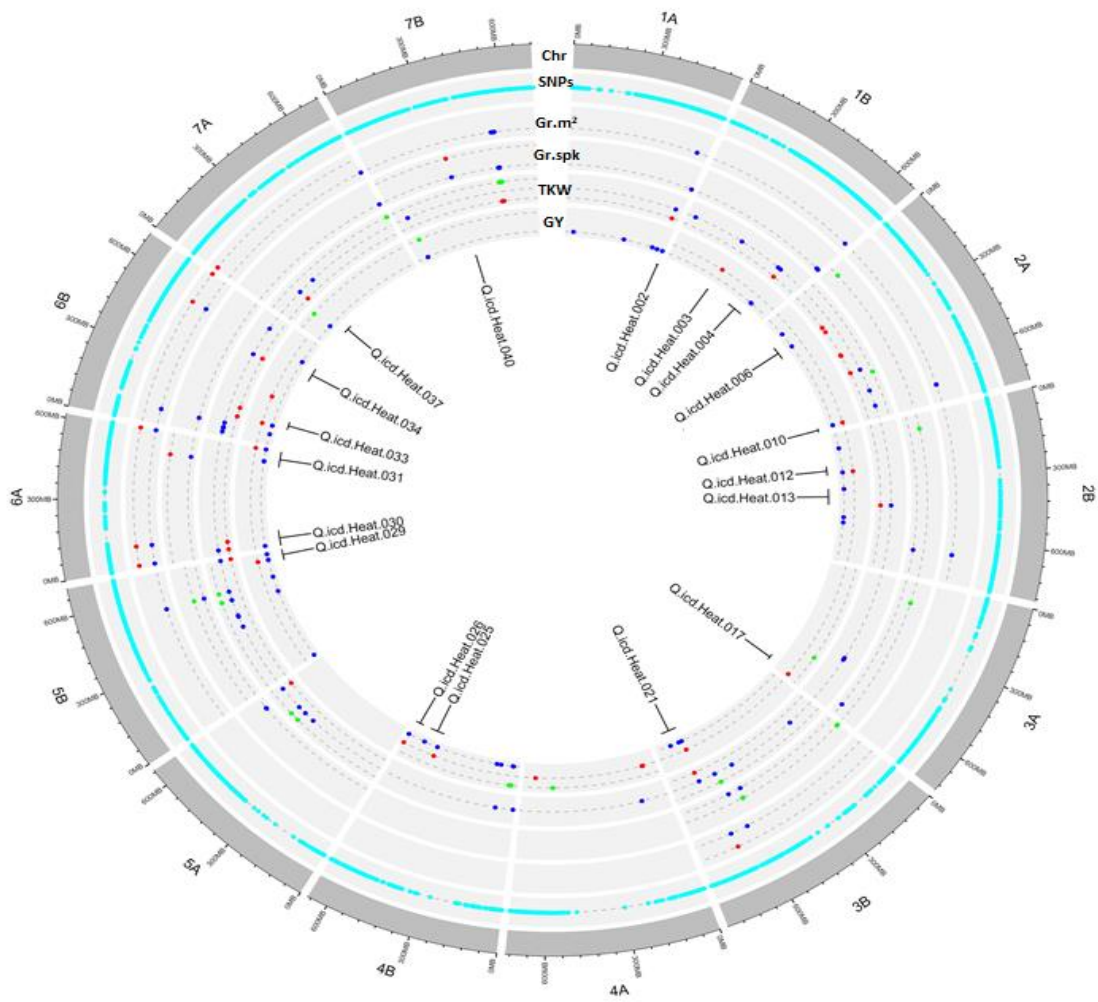

Figure III- 5: Circoss representation of markers-trait association for the traits GY, TKW, Gr.m-2 and Gr.spk and their HSI and the identified QTLs under heat stress.

III-3-4. Markers validation by KASP

Markers validation is the first step required before deploying the most interesting QTLs via MAS, for that a total of 20 MTA sequences linked to important agronomical and spike fertility traits were submitted for KASP primers design. Among these, only 14 could be successfully designed, and 11 identified a polymorphism within the validation set. Three showed significant ($p < 0.05$) correlation to the test phenotype. Three QTLs were represented by these three markers. AX-95632723 tagged Q.icd.Heat.09 on chromosome 3A, AX-94932858 tagged Q.icd.Heat.021 on chromosome 3B and AX-94588421 underlines Q.icd.Heat.029 on chromosome 5B. AX-95632723 reached 5% correlation to grain yield under heat,65% sensitivity, 73% accuracy and 80% specificity. AX-94932858 also reached significant correlation of 13% to HSI-GY, Gr.m-2 and HIS-Gr.m-2 under severe heat, 100% sensitivity, 55% accuracy and 10% specificity. However, AX-94588421 reached only 4% of correlation to HSI-GY. 100% sensitivity, 60% accuracy and 20% specificity. Overall, AX-94932858 appeared as the most suitable for MAS application.

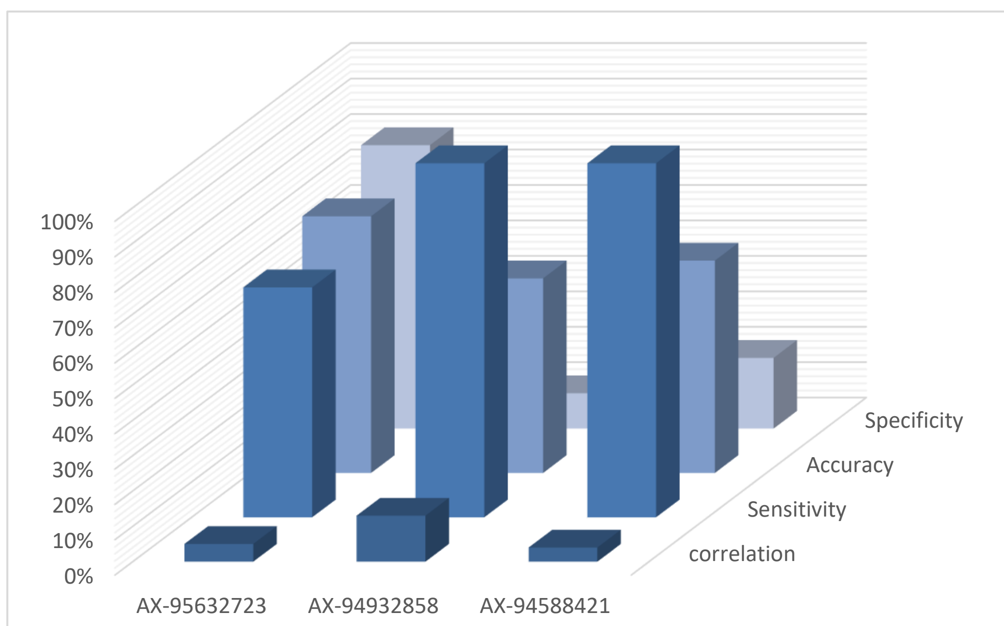

Figure III- 6: Kompetitive Allele Specific PCR (KASP) markers validation on an independent set of 94 elite lines of ICARDA tested under severe heat for grain yield and biomass. Correlation was measured between the BLUE for grain yield recorded along the Senegal River and the haplotype score. Accuracy, sensitivity, and specificity where determined using only the top 20 and worst 20 lines.

Table III- 4 : Quantitative trait-locus (QTL) for yield component traits and their heat susceptibility index

(“” indicate the presence of the QTL in which cluster, and for the HSI of the trait)*

Locus	Key marker	Chr	Position	Interval size	Max LOD	Max r2	Trait	Stress env	Normal env	HSI
Q.icd.Heat.002	AX-94503114	1A	569666989	10937990	6.46	18%	GY, Gr.spk	*		*
Q.icd.Heat.003	AX-95097722	1B	303951517	5280568	3.24	9%	GY,TKW	*		*
Q.icd.Heat.004	AX-94583358	1B	464812402	97859317	4.34	10%	GY,TKW	*		*
Q.icd.Heat.005	AX-94530424	2A	43197054	70934610	4.11	9%	GY, Gr.spk	*	*	*
Q.icd.Heat.006	AX-95179814	2A	156411002	83933730	6.41	15%	GY,TKW	*		*
Q.icd.Heat.010	AX-94754539	2A	748712624	20911329	6.8	15%	GY	*		*
Q.icd.Heat.011	AX-95019864	2B	69780105	62510644	4.11	9%	GY, Gr.spk	*	*	*
Q.icd.Heat.012	AX-94735056	2B	199934683	55764396	7.19	15%	GY	*		*
Q.icd.Heat.013	AX-94971410	2B	347564919	97653476	4.61	9%	GY,TKW	*		*
Q.icd.Heat.017	AX-95024816	3A	685995771	36068974	4.29	8%	GY, Gr.spk	*		*
Q.icd.Heat.021	AX-94932858	3B	672663491	90927353	6.87	16%	GY,Gr.m-^{2}, TKW	*		*
Q.icd.Heat.024	AX-94539986	4B	6902167	99194545	5.68	16%	GY,TKW	*	*	*
Q.icd.Heat.025	AX-94496513	4B	484213065	83324043	3.78	6%	GY	*		*
Q.icd.Heat.026	AX-94912895	4B	620613828	47470793	3.67	8%	GY	*		*
Q.icd.Heat.029	AX-94559013	5B	634496056	66468163	5.26	12%	GY,TKW	*		*
Q.icd.Heat.030	AX-94560148	6A	1035225	83517466	6.45	15%	GY,Gr.m-^{2}, TKW	*		*
Q.icd.Heat.031	AX-94562707	6A	512570339	99435766	6.37	17%	GY,Gr.m-^{2},Gr.spk	*		*
Q.icd.Heat.033	AX-94994562	6B	121497853	48753486	3.88	9%	GY,TKW	*		*
Q.icd.Heat.034	AX-94403103	6B	463748670	87655142	6.15	16%	GY,Gr.m-^{2}, TKW	*		*
Q.icd.Heat.036	AX-94606531	7A	9137789	94445380	3.92	9%	GY,Gr.m-^{2}	*	*	*
Q.icd.Heat.037	AX-94483534	7A	130624863	53188872	3.83	10%	TKW,Gr.m-^{2}	*	*	*
Q.icd.Heat.040	AX-95125910	7B	327094702	10	10.47	26%	Gr.spk	*		*

III-4. Discussion

III-4-1. Performances of the panel under heat-stressed conditions

This study assessed 216 durum wheat modern lines from nine countries of origin including ICARDA and CIMMT lines. These were planted in the South of Morocco (Tassaout and Melk zhar) and along the Senegal River (Fanaye and kaedi) to test their ability to withstand extreme heat throughout the growing season. There was a 24º-latitude difference between Moroccan stations and Senegal River stations. While all four stations experienced a significant heat stress throughout the growing cycle, the heat stress was less intense in Tessaout and Melk Zehr with maximum temperature around at flowering 26.6 and 30.4 °C during the second crop seasons, respectively. Fanaye and Kaedi were much warmer (maximum temperature: 37.1 and 39 °C during the second crop season, respectively) (Annex 2).

Agronomic performances for GY, TKW, and Gr.spk were recorded to reveal that Fanaye's mean yield was reduced by 35% and 52% compared to Melk Zhar during 2015 and 2016, respectively, and by 60% compared to Tessaout in 2016. While in Kaedi the mean yield was reduced by 52% and 46% compared to Melk Zhar (during 2015 and 2016, respectively) and by 55% compared to Tessaout in 2016. Similar results were reported previously by Sukumaran et al. (2018), who indicated that mean yield for CIMMYT durum wheat lines was reduced by 72% when the genotypes were grown under heat stress conditions.

The most surprising result was the excellent top yield achieved along the Senegal River, with productivities of up to 6.6 t ha^{-1} in Fanaye and 4.7 t ha^{-1} in Kaedi with seasons of just 92 days. In fact, the large kernels obtained along the River (47g in Kaedi and 49 g in Fanaye) allowed reaching high productivity even under a severe temperature stress. Also, Kumar et al. 2013 indicated that thousand kernel weight (TKW) has very important contributing against heat stress. However, Sall et al. (2018, 2019) failed to reveal this trait as critical for heat tolerance and rather indicated that grain setting (Gr.spk) was the most important. This is likely due to the fact that only heat stressed environments were used in that study, and the lack of contrasting normal conditions failed to reveal the importance that this trait might cover. For this purpose, a heat susceptibility index (HSI) was estimated to compare between normal and stressed sites. Heat tolerance is a trait quantitative in nature, controlled by a large number of genes, which may be involved in interactions with each other (QTL × QTL interaction) and/or with the environments (QTL × E and QTL × QTL × E interactions) (Kumar et al., 2013). To dissect theses interactions, three environmental clusters were

identified based on weather and phenological data and used to estimate HIS corrected for GxE effects. While most of the studies that have been reported thus far used HSI under late-sown heat-stressed conditions and controlled conditions in soybean and wheat (Ayeneh et al., 2002; Githiri et al., 2006; Kirigwi et al., 2007; Mohammadi et al., 2008; Pinto et al., 2010; Yang et al., 2010; Barakat et al., 2011; Mason et al., 2010, 2011), here actual heat-stressed environment was used to derive the contrasting value. Therefore, it is suggested by Mason et al. (2010) that HSI may be used as an indicator of yield stability and a proxy for heat tolerance.

The results indicated that 40% of the lines (Figure III-4) showed good performance and were considered as heat tolerant. In particular, IDON37-013, IDON37-081 and IDON37-097 were the top ones and were considered as highly tolerant for all four traits combined. These should be kept in high consideration for deployment in breeding programs.

III-4-2. Genomic regions controlling heat tolerance in durum wheat

GWAS is the most popular approach for dissecting the genetic basis of complex traits (Sukumaran et al., 2015), but this approach is prone to the detection of false positives due to confounding population structure or the effect of phenology genes (Yu et al., 2006; Zhang et al., 2010). Our study used the mixed model framework of Yu et al. (2006), with fixed and random effects to control false positives. Q-Q plots for multiple models were evaluated to select the best models for identifying MTAs (Sukumaran et al., 2015), mixed linear model MLM was deemed the best model for identifying MTAs. To avoid noise due to phenology, days to heading (DTH) was used as covariate, and Bonferroni correction was used to avoid the identification of spurious associations. Finally, the stringent protocol followed here recognized as true MTAs only those loci identified both by individual trait analysis as well as HSI.

A total number of 63 MTAs were detected for individual traits among clusters, TKW had the highest number of MTAs (46) followed by Gr.spk (10), while GY and Gr.m^{-2} had the lowest number of MTAs (6 and 1 respectively). When, 157 MTAs were detected for the heat susceptibility index of the traits with the following proportions: 54 MTAs for HSI-GY, 43 for HSI-TKW, 33 for HSI- Gr.m^{-2} and 21 for HSI-Gr.spk.

29 MTAs were detected in common for the individual traits and their HSI and were considered as highly significant most of them were associated with TKW and Gr.m^{-2} . The individual traits TKW and Gr.m^{-2} can be manipulated, independently of grain yield, since unique MTAs were observed for them under different conditions (Sukumaran et al., 2018). Co-localizing MTAs were identified for GY, TKW, Gr.m^{-2} and Gr.spk, and

individual MTAs were observed for each trait. While, the phenotypic variation explained by the markers was ranged from 1 to 18% along with LOD ranged from 3.3 to 9. This result suggests complementing the SNP-based GWAS by haplotype based GWAS to explain the missing variation (N'Diaye et al., 2017).

The MTAs were located in 25 genomic regions (QTLs). Out of 25, eight QTLs were observed as unique for a trait, for GY (4) in chromosomes 1B, 3A, 4A and 7B, and for TKW (4) in chromosomes 2B, 3A, 5A, 5B, 7A and 7B.

Where, 17 QTLs were considered as multi-trait. Among these, 9 QTLs were identified in cluster1 under non-stressed conditions, meaning that their involvement, even though of great importance, it is not directly linked to heat tolerance. In particular, Q.icd.Heat.02 on chromosome 1A, had a max LOD of 6.46 it was involved in controlling the four traits and the HSI of GY, Gr.m^{-2} and Gr.spk, and it explained 18% of phenotypic variation. While, in the recent study of Sukumaran et al. (2018) on durum wheat CIMMYT lines under heat stress, no such QTL effect was reported.

seven major multiple traits QTLs were detected on chromosomes 1A,2A, 4A, 4B, 5A, 5B and 7A with LODs ranged from 3.4 and 7.6 and explained phenotypic variation between 8 and 18%. They were expressed in the three clusters, regardless of heat stress pressure.

Q.icd.Heat.14 was considered as heat-related QTLs and was identified in chromosome 4B and linked to the control of GY and the HSI of GY, TKW and Gr.m^{-2}. This result agrees with the finding of Sukumaran et al. (2018) that also identified a genomic region associated to HSI-GY in the same chromosome under heat pressure.

The co-location of QTL for different agronomic and physiological traits with QTL for yield suggests that it is possible to achieve genetic dissection of the crop performance under heat stress to facilitate a more efficient breeding approach (Pinto et al., 2010).

On chromosome 7B, a highly significant QTL ("Q.icd.Heat.40", LOD=10.47 and r^2=26%), was detected and covered the association to Gr.spk and its HSI. previously, Paliwal et al.,2012, presented a similar QTL effect that control the HSI of GY and TKW on long arm of the same chromosome in hexaploid wheat under late sown conditions. Also, within the same genomic region, Sukumaran et al. (2018) identified QTL that control tolerance index of the TKW on durum wheat CIMMYT lines under heat stress. These regions should be kept in high consideration and go for marker validation.

Furthermore, a KASP marker has already been validated to help accelerate the selection for the most important QTL identified here (Q.icd.Heat.021). with significant accuracy showed, it was possible to use it to select for the positive allele in a validation

set of completely independent germplasm with different genetic backgrounds. As such, we believe its exploitation in breeding can begin with confidence until a better marker can be develop. The remaining QTLs will also need to be converted into KASP markers, but their validation will have to wait until ad hoc populations integrating different parents are devised and tested under severe heat stress.

III-5. Conclusions

The results presented here provide an ideal opportunity to conduct pyramiding via molecular breeding by MAS (Yang et al., 2010). Traits like GY, Gr.spk, Gr.m^{-2}and TKW are significantly affected by the environment, making their selection difficult to achieve without errors. Therefore, the molecular markers associated to these traits identified in this study would provide an ideal solution to avoid the confusing effect of the environment. Three genotypes have been identified as extremely heat tolerant, IDON37-013, IDON37-081 and IDON37-097 with each carrying 91%, 88% and 89% of positive QTLs respectively. At least, 4 QTLs are present in one but not all three of the genotypes, indicating a good opportunity to pyramid positive alleles together. First, a validation step will be required to convert the expensive Axiom markers into easier to use types. Afterward, MAS could be deployed to accelerate the breeding process. Also, genomic selection could be deployed to capture not only the major but also the minor useful alleles that could not be easily detected via GWAS (Bassi et al., 2016). The combination of phenotypic superior lines and molecular tools to accelerate selection, should truly enable durum wheat breeders to rapidly achieve new heat tolerant cultivars ready to face the climatic challenges of tomorrow.

Chapter IV: Combining QTL Analysis and Genomic Predictions for Four Durum Wheat Populations Under Drought Conditions

Hafssa Kabbaj[1,2], Meryem zaim[2], Zakaria Kehel[1], Gregor Gorjan[4], Abdelkarim Filali-Maltouf[2], Bouchra Belkadi[2], Rodomiro Ortiz [3], Filippo M. Bassi[1]*

1 International Center for Agricultural Research in the Dry Areas, Rabat, Morocco,
2 Department of Plant Science, Mohammed V University, Rabat, Morocco,
3 Plant Breeding, Swedish University of Agricultural Sciences, Alnarp, Sweden
4 Zootechnical Department, Biotechnical Faculty, University of Ljubljana, Domzale, Slovenia

Abstract

Durum wheat is an important crop for the human diet, and it is consumed largely in the form of traditional dishes such as pasta, couscous, unraised breads, and bourghul. Worldwide, the consumption of durum wheat is gaining popularity because of its nutritional properties. In order to ensure that durum wheat production maintains the pace with the increase in demand, it is necessary an increase in production of approximately 1.5% per year. Genomic selection (GS) recently appeared as an ideal tool to deliver this level of annual genetic gain, but its deployment in durum wheat breeding remains extremely limited. Here, we present a pilot study aimed at refining the mode of potential deployment of GS in durum wheat breeding. A total of 576 individuals were sown at three locations in Morocco and one in Lebanon. These individuals were genotyped by sequencing with 3,202 high-confidence polymorphic markers and were used to train statistical model incorporating relationship matrix, the genotype by environment interaction, and marker by environment interaction to reveal significant advantages for models incorporating the marker effect. Using training populations (TP) in full sibs relationships with the validation population (VP) was shown to be the only effective strategy, with accuracies reaching 0.35-0.47 for GY. Reducing the number of markers to 10% of the whole set, and the TP size to 20% resulted in non-significant changes in accuracies. The QTL identified were also incorporated in the models as fixed effect, showing significant accuracy gain for all four populations. Our results confirm that the prediction accuracy depend greatly on the relatedness between TP and VP, but not on the number of markers and size of TP used. Further, feeding the model with information on markers associated with QTLs increased the overall accuracy.

IV-1. Introduction

Despite important strides in marker technologies, the deployment of marker-assisted selection (MAS) has failed to provide significant improvements in agronomic performances for quantitative traits. Current MAS methods are better suited for manipulating a few major effect genes (Dekkers and Hospital, 2002). Unfortunately, the small-effect genes underlie most of the complex polygenic traits that are crucial for the success of new crop varieties (Crosbie et al., 2003). On the other hand, the method defined as genomic selection (GS) proposed by Meuwissen et al. (2001) presents several advantages in handling these minor effect alleles and could provide a strategic solution to accelerate the gentic gain in wheat breeding programs (Bassi et al. 2017).

Genomic selection (GS) analyzes jointly all markers on a population, in an attempt to explain the total genetic variance through the sum of the markers effects (Meuwissen et al., 2001). The 'genomic estimated breeding value' (GEBVs; Meuwissen et al., 2001) is calculated for individuals for which only the genotypic information are available, by modeling individuals for which phenotypic and genotypic data are collected. The set of individuals with available phenotypic and genotypic data is defined as the 'training population' (TP), while the set of individuals which are only genotyped, and from which the selection is made, is defined as the 'breeding population' (BP) (Figure IV-1). The 'accuracy' of the predicted GEBV is determined by the correlation between GEBV and the true breeding value (TBV) calculated phenotypically for a 'validation population' (VP), which is genotyped and phenotyped but not used to train the model. The value for accuracy is often used to determine the statistical model that best fit the population under study, and also to assess the overall success of the GS approach. Therefore, it is important to maintain a high degree of accuracy, and hence to use a TP that best fits the BP. The degree of relatedness between the two populations is often a good predictor of the accuracy that will be achieved. Cross-validation is used to train and develop the prediction models using different sampling techniques in the training population data sets ahead of estimating the GEBVs in the BP. The idea behind this approach is that breeders can derive predictions of the breeding value of an experimental line even before the line has been tested in the field. In turns, this would allow to take decisions on the use of the lines for yield testing or crossing already during the earlier generations. Therefore, it can reduce breeding cycles time from 9 to 3 years, increasing by 2-3 folds the rate of annual genetic gain per unit of time.

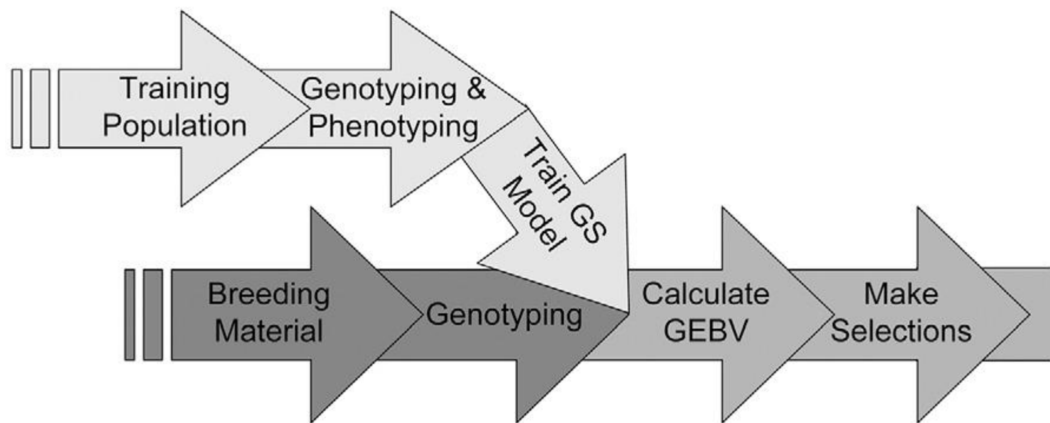

Figure IV- 1: Diagram of genomic selection (GS) processes starting from the training population and selection candidates continuing through to genomic estimated breeding value (GEBV)–based selection. (Heffiner et al.,2009)

Progress in the field of medium- and high-throughput genotyping platforms alongside decreased costs for marker detection via sequencing technologies has enabled the application of GS (Davey et al., 2011). Several authors have used breeding data with individuals genotyped with intermediate- to high-density markers to show that traits such as grain yield, biomass yield, disease resistance and flowering evaluated under different environmental conditions can be predicted with varying levels of accuracy depending on, among other factors, the heritability of the trait, the size of the training population, the number of markers, the relationship between the training and testing sets, and genotype by environment interaction (GE) (de los Campos et al., 2009; Crossa et al., 2010, 2011; Pérez et al., 2012; Burguen˜o et al., 2012; Gonzalez-Camacho et al., 2012; Hickey et al., 2012; Riedelsheimer et al., 2012; Windhausen et al., 2012; Zhao et al., 2013).

Markers used for GS must be able to tag all loci that explain some of the phenotypic variation of the trait of interest in the selection population. Implying that linkage disequilibrium (LD) between neighboring markers and accuracy of genotyping must be high (Jonas and Koning, 2013). Moreover, the extent of LD has to be considered because they can differ greatly between breeding populations and have large impacts on the design of genotyping platforms (Hamblin, M.T. et al. 2011; Wientjes, Y.C.J. et al. 2013). Higher marker densities are required if LD is expected to be low, such as in bi-parental populations, which are common in crop breeding (Asoro, F.G. et al. 2011). Additional challenges have to be overcome before GS can be fully implemented for the selection of diverse (crop) species such as the application in biparental breeding and

(multiple) crossbred populations, coverage of complex genome structures, capture of effects caused by epistasis, genotype by environment interaction (GE) and the effect of rare alleles (Heffner, E.L. et al., 2011a).

The variety and complexity of crop breeding cycles requires in-depth studies on if, how, and when to apply GS. In the case of durum wheat, several questions remain unanswered but some clarity has become to appear. Firstly, the use of the different statistical models available thus far does not appear to cause major changes in accuracy (Hayes et al., 2009; Zhong et al., 2009; VanRaden, 2008; Lorenzana and Bernardo 2009). In particular, the Bayesian Ridge Regression (BRR) has become the model of choice thanks to its low computational cost and higher accuracy. Secondly, the number of markers to be used in wheat, while it might vary based on the level of inbreeding to be characterized (Bassi et al. 2016) and the population structure, it appears to plateau in accuracy gains at approximately 500 polymorphic loci (Heffner, et al., 2011b; Hickey et al., 2014). Third, model that do not account for GxE, hence for which multiple environment data are not available, are incapable of predicting low heritability traits such as grain yield (Bassi et al. 2016). Fourth and most important, the type of TP to be used remains an argument for discussion. While it is clear that the level of relatedness between TP and BP/VP is of the greatest importance, the number of studies comparing weakly related, half, and full sibs is still very low. Finally, the use of background makers in combination with major known QTLs has been suggested but never actually measured.

Here, four recombinant inbreed lines (RILs) of durum wheat with different level of relatedness were tested across environments, to attempt to answer some of the questions listed above.

IV-2. Materials and methods

IV-2-1. Plant material

Four RILs populations were used to study the application of GS in durum wheat: Jennah Khetifa/Cham1/3/Omrabi5/T.dicoccoides600545//Omrabi5 (DRO) developed for pyramiding drought tolerance genes ; SW-Algia/Gidara1/Cham1 (SW) for Septoria tritici resistance; Younes/Gidara2 (YG) for combining drought tolerance and yield potential; and Icamor/Gidara2 (IG) for combining Hessian fly resistance with high yield potential. A total of 576 individuals were sown at three locations in Morocco Jemaa Shaim (JSH; 32°21′0″N and 8°51′0″W), Marchouch (MCH; 33°34'3.1"N and 6°38'0.1"W) and Sidi el Aidi (SAD; 33°9'36"N and 7°24'0"W), and one station in

Lebanon, Terbol (TER; 33°48'29"N and 35°59'22"W). All population were planted in plots of 6 m² at a seeding rate of 120 kg ha^{-1} using augmented randomized complete block design with repeated checks. YG (145 individuals) was planted in MCH, JSH, SAD and TER. DRO (197 individuals) population was also planted in all stations except TER. IG (115) was sown in MCH and TER. While 93 lines of SW were present only in MCH. 1,000 kernel weight (g) (TKW) and grain yield (kg ha-1) (GY) were recorded in all locations.

IV-2-2. DNA extraction and genotyping

Leaf samples obtained from F9 plants were freeze-dried and used for C-TAB DNA extraction. DNA quality was assessed on agarose gel and it was then equilibrated to 100 ng. The DNA was shipped to the Poland lab at Kansas State University for genotyping by sequencing following the protocol of Poland et al. 2012. Briefly, two restriction enzymes (*PstI* and *MspI*) were used for genome complexity reduction, followed by 96-multiplex sequencing by barcoding. Low-quality data filtering was carried out according to the following rules: heterozygous calls not superior to 2%, maximum of 30% missing data, and a minor allele frequency superior to 10%.

IV-2-3. Statistical analysis

Grain yield (GY) and 1,000 kernel weight (TKW) were first adjust for field heterogeneity using board sense heritability (H^2) that was calculated for each pop within and across environments. The best linear unbiased estimator (BLUEs) were then estimated for all population and environments.

A total of four genomic models were tested as first step in this study:

i) a baseline additive model without interactions of genotypic effect (G), environmental (E) effect, and error (ε) (G+E+ ε)

ii) a baseline multi-environment model (G+E+GxE+ ε), which assumed interactions between the G and the E.

In both these models, all the effect were assumed to be random with a normal distribution N(0, σ) where σ is the term variance

iii) the third model was a marker (M) effect model (G+E+GxE+M+ε), where the genotype effect is substituted by an approximation of the genotype's genomic value expressed as a regression on marker covariates.

In this case the model assumes that the genotype's genomic value follows a normal distribution N(0, G σg) where σg is the genetic variance and G is genomic relationship matrix.

iv) the last model is the marker x environment model (G+E+GxE+M+MxE+ ε) where the marker effect is composed by an effect common to all environment (main effect) plus a random deviation specific to a particular environment (López-Cruz et al. 2015).

Testing of the different models accuracies was done using DRO, IG and YG populations independently, and setting as cross-validation 80% of the individuals as TP and 20% as VP. The accuracies within and across environments were then calculated as a measure of good fit. The BGLR package (Pérez & de Los Campos, 2014) was used to run all models above from (i) to (vii) by Bayesian ridge regression (BRR) using 10,000 iterations and 5,000 burn in, and 50 replications (de los Campos et al., 2009; 2013). This model induces homogeneous shrinkage of all marker effects towards zero and yields a Gaussian distribution of marker effects. The 50 replications were used to define statistical differences between model accuracies following a one factor ANOVA.

The GxE + MxE model (iv) was selected and used to test additional hypothesis:

v) the effect of markers number was investigated by comparing predictions using 100%, 80%, 60%, 40%, 20% and 10% of the total marker set in combination with reducing the TP population size to 20%, 50%, and 75% for GY and TKW. The TP individuals were selected randomly in 50 replications, and one factor ANOVA was used to determine significant differences.

vi) the prediction accuracy of using half sibs vs full sibs as TP were compared. Each population was set as TP for all others and itself using the whole population as TP and the whole other population as VP.

vii) to compare the value of MAS and GS, the prediction accuracy was calculated using 50% as TP and 50% as VP for all markers, only markers associated with major effect QTLs, with 44 and 27 markers for GY and TKW, respectively, and by removing these markers linked to QTLs from the set. The TP individuals were selected randomly in 50 replications, and one factor ANOVA was used to determine significant differences.

viii) the rr-BLUP package v4.6 (Endelman, 2011) was used to run a mixed model estimating the accuracy gain when using markers underlying the QTLs as fixed effects, and the remaining markers as random effects. For this analysis ten random subsets of 50% TP and 50% VP were selected in each population separately (DRO, IG, SW and YG). QTL analysis was conducted again for each TP subset following the method described above. Those

markers that resulted as underlying QTLs in each TP subset were fixed in the model. One factor ANOVA was run for the ten replicates of each population to determine significant differences.

IV-3. Results

Four recombinant inbreed lines (RILs) of durum wheat were used to train genomic prediction different statistical models. The board sense heritability was calculated for GY and TKW in each population, the results presented in Table IV-1 show that the heritability for GY was above 0.5 in Marchouch and Sidi ayedi for DRO population and in Terbol for IG population, and it ranged between 0.1 to 0.5 for TKW. Overall the heritability was high (more than 0.5) except for YG population in Terbol where it was equal to 0.2.

Table IV- 1: Board sense heritability (H2) for grain yield (GY) and 1,000 kernel weight (TKW) for the four population (DRO, IG, YG, SW) across environments.

		Environnements			
Population	**Trait**	**Marchouch**	**Jemaat shem**	**Sidi ayedi**	**Terbol**
DRO	GY	0.6	0.3	0.8	X
	TKW	0.8	0.6	X	X
IG	GY	0.1	X	X	0.8
	TKW	0.5	X	X	0.5
YG	GY	0.1	0.2	0.4	0.5
	TKW	0.7	0.4		0.2
SW	GY	0.3	X	X	X
	TKW	0.7	X	X	X

IV-3-1. Genomic prediction: identification of the best fitting model (i, ii, iii, iv)

Four statistical models (i, ii, iii, iv) were tested to determine the best model to be used for each population (Figure IV-2). Non-significant differences could be identified for the IG population with average accuracies that ranged from 0.42 to 0.41. For DRO, the incorporation of the M effect resulted in a significant increase in accuracy from 0.47 to 0.49. The YG population was the most sensitive to the change of model ranging from 0.27 for models without M (i and ii), to 0.30 for model iii, to 0.33 for model iv incorporating GxE+MxE. According with these results, the model incorporating GxE +

MxE was chosen to be the best fitting for all three populations. For SW population phenotypic data were available only for one environment, therefore a model using only markers effect (iii) was used to run genomic predictions for this population

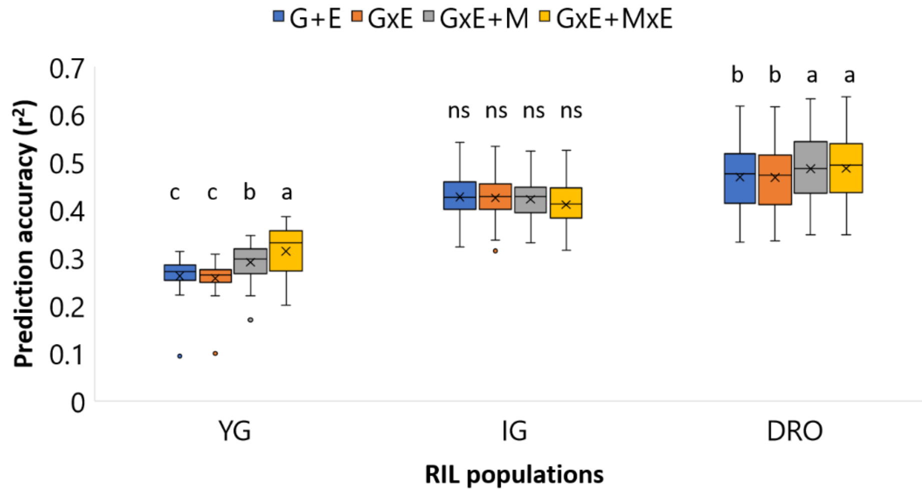

Figure IV- 2 : Prediction accuracy for grain yield (GY) in YG, IG and DRO populations using four different statistical models. G+E: genotype + environment effect, GxE: genotype by environment interaction, GxE+M: genotype by environment interaction +markers effect; GxE+MxE: genotype by environment interaction + markers by environment interaction. The horizontal line represents the average, the square indicates the 2nd and 3rd quartiles, the whiskers represent the 1st and 4th quartiles, the cross the median, and the dots are outliers. The letters indicated classes determined via LSD.

IV-3-2. Genomic prediction: effect of reducing TP and marker size (v)

The effect of marker number and TP size on prediction accuracies was tested for GY and TKW (v). Figure IV-3 shows that when decreasing the number of markers from 3,202 to 320, a slight decrease in prediction accuracies was observed for the different set of TP. For GY, the reduction of markers caused a shift from 0.44 to 0.41 accuracy using 20% of TP, from 0.47 to 0.43 and from 0.49 to 0.44 for 50% and 75% of TP, respectively. For TKW, it dropped from 0.75 to 0.73 and from 0.76 to 0.74 for 20% and 50% of the TP, respectively, while no difference was observed for the 75% of TP between the total number of marker and 10% of it. Statistical analysis revealed no significant differences could be observed when reducing marker number and TP size for any of the two traits.

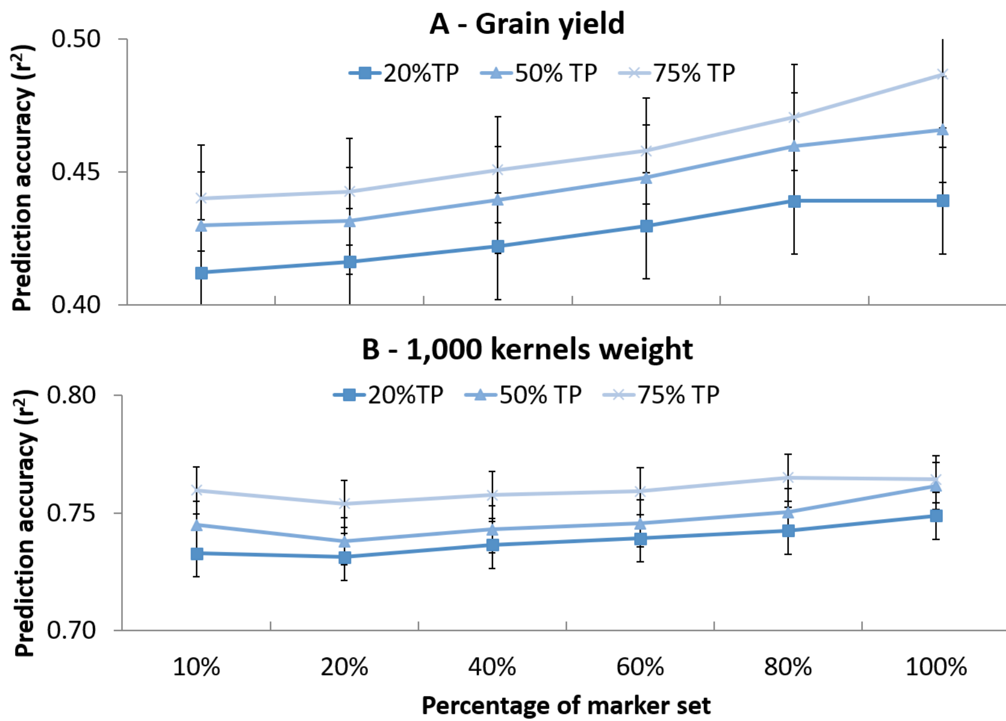

Figure IV- 3: Prediction accuracy for grain yield (A) and 1,000 kernel weight (B) using different randomly selected sub-sets of markers in decreasing order: 320 (10%), 640 (20%), 1,281 (40%), 921 (60%), 2,562 (80%), and 3,202 (100%) tested on DRO population using 20%, 50% and 75% of the whole population as training set (TP) to predict the rest of the population (VP). The whiskers represent the standard errors.

IV-3-3. Genomic prediction: importance of relatedness between TP and VP (vi)

The four populations share common parents and have hence kinship relationships. It was therefore evaluated if it would be possible to use one population as TP for the others (VP) which have half-sibs relationships. Using TP that were full sibs to the VP resulted in good accuracy values that ranged from 0.35 to 0.47, and from 0.92 to 0.30 for GY and TKW, respectively (Table IV-2). When the TP was not derived from the same cross of the BP (half sibs), the accuracies drop to values close to zero or even negative (Table IV-2). The only acceptable case for GY with accuracy of 0.29 was obtained when SW was used as TP for IG, but this was not true when IG was used as TP for SW (accuracy of 0.08). The same was observed for TKW, with SW as TP ensuring accuracy of 0.22, while YG as TP dropped to 0.09 accuracy. Interestingly, the two most genetically related populations IG and DRO also resulted in very poor prediction accuracies when used as TP for each other.

Table IV- 2. Comparison of the prediction accuracies using full sibs and half sibs as training populations for grain yield and 1,000 kernel weight. The columns represent the TP and the rows are the BP, the diagonal represents the full sibs relationships

	DRO	IG	YG	SW	DRO	IG	YG	SW
	Grain yield				1,000-kernels weight			
DRO	0.47	-0.08	-0.11	0.07	0.76	-0.1	0.03	-0.26
IG	-0.09	0.41	0	0.08	-0.08	0.92	-0.02	0.09
YG	-0.07	-0.02	0.35	-0.08	0.12	0	0.83	0.14
SW	0.06	0.29	-0.13	0.37	-0.26	0.22	0.11	0.3

IV-3-4. Genomic prediction: effect of QTL analysis on model accuracy (vi, viii)

Since QTL analysis and GS have been rarely combined, the last objective of this study was to determine if a step of QTL analysis could help improve the GS model accuracy. A total of 44 and 27 markers were associated via QTL analysis to GY and TKW, respectively. To test their value alone, these were used as the only marker to perform genomic predictions and resulted in non-significant accuracies for GY for DRO (0.18), and IG (-0.02), while significant accuracies could be identified for YG (0.29), while an increased was observed for SW (0.54). Similarly, for TKW there was a loss significance for DRO (0.20), IG (0.11) and YG (0.09), while it again increased for SW (0.54) (Figure 4). The opposite attempt was also conducted by removing from the whole set al.l the markers associated with QTLs. In this case the GY and TKW accuracies became non-significant for all populations, except for SW for which it matched what obtained when using the full marker dataset (Figure IV-4). With the exception of SW, for which the use of only markers associated to QTLs had a positive effect on the prediction accuracies, in all other populations the use of all markers combined was significantly superior.

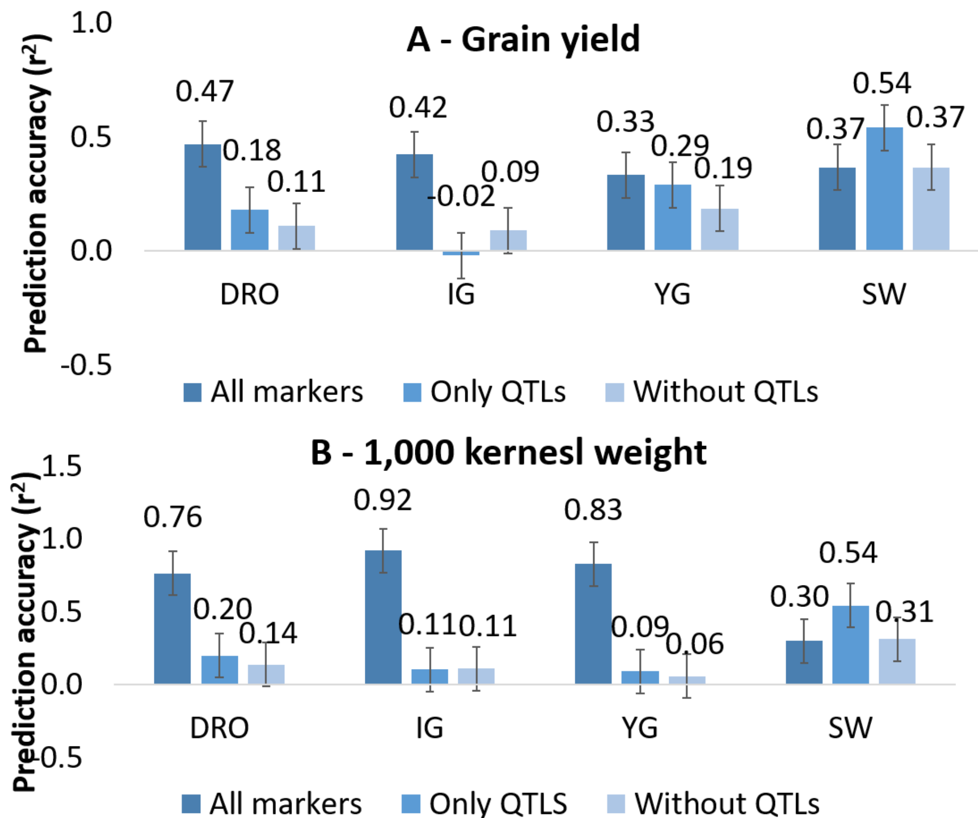

Figure IV- 4: Prediction accuracy for grain yield (A) and 1,000 kernel weight (B) using all markers, only markers linked to QTLs, and all markers except those identified as linked to QTLs. Whiskers represents the experiment wise LSD.

As it can be expected, the sum of the accuracies of using markers associated to large and small effects does not equal to the accuracy of these combined. It then becomes interesting to assess a model that better incorporates these two by fixing the effect of markers associated to QTLs, while including the random effect of the small impact alleles (viii). To test the suitability to do so in a context that better resemble an actual breeding pipeline, QTL discovery was re-run for each random group of entries composing the TP, and only QTL that could be identified by the specific TP where fixed in the model. In annex 4 is reported how frequently the QTL associated with GY could be re-identified for each TP sub-set. The results of fixing the marker underlying the QTLs in the model is reported in Figure IV-5. For all four populations the accuracies increased significantly ($p<0.05$) when the QTL-underlying markers were fixed in the model. The average accuracies shifted from 0.35 to 0.47, 0.38 to 0.44, 0.29 to 0.35, and 0.35 to 0.41, for YG, DRO, IG, and SW populations, respectively. This represents a clear gain of 0.06 to 0.12 points of accuracy, superior than the 0.01 to 0.03 obtained by testing different statistical models (i, ii, iii, iv).

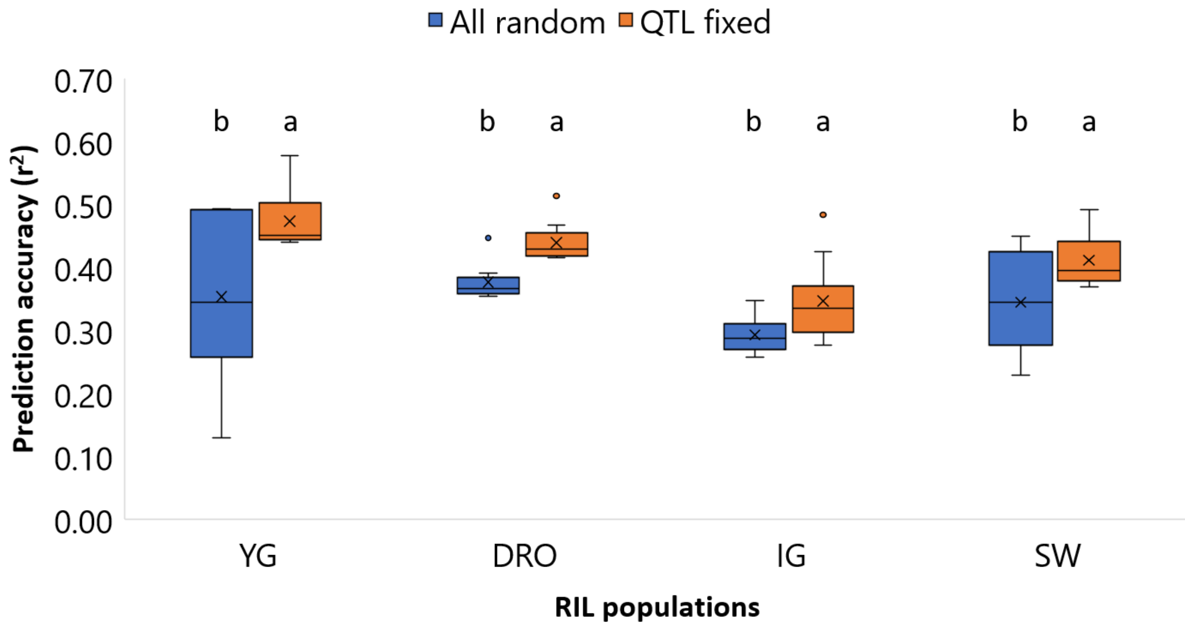

Figure IV- 5: Comparison of the prediction accuracies of grain yield (GY) for the four population YG, DRO, IG, and SW, using a model with all markers considered as random effect against a models that fixed markers underlying QTLs. The horizontal line represents the average, the square indicates the 2nd and 3rd quartiles, the whiskers represent the 1st and 4th quartiles, the cross the median, and the dots are outliers. The letters indicated classes determined via LSD.

IV-4. Discussion

Genomic selection is a form of MAS that simultaneously estimates all locus, haplotype, or marker effects across the entire genome to calculate genomic estimated breeding values (GEBVs; Meuwissen et al., 2001).

Genomic predicting was estimated in four RILs populations of durum wheat, planted in four environments during the season 2014/2015. The total of 546 individuals were genotyped with ‘genotyping by sequencing’ (GBS) method, a consensus map was constructed, and imputation was performed to generate a total of 3,202 polymorphic SNPs markers to conduct this study. Poland et al., 2012 indicated that no significant differences were observed in the accuracy of genomic-estimated breeding values (GEBVs) among imputation methods, where the values of prediction accuracies with GBS were 0.28 to 0.45 for grain yield. Hence, the use of imputation method in this study should not have an effect on the overall quality of the analysis. The Bayesian ridge regression statistical model was used to carry out the genomic prediction for two complex traits: GY and TKW. This model approach is more appropriate than shrinkage

regression in which there are few or no large effects and many small effects (Breiman, 1995), as is the case with most quantitative traits.

Board sense heritability for grain yield (GY) ranged from 0.10 to 0.85 depending on the population and environment condition, while it was between 0.22 and 0.81 (Table IV-1) for TKW. Low heritability traits pose a problem for both phenotypic and genomic selection, and thus require larger TP and number of test locations in order to maintain high accuracies. Hayes et al., 2009 indicated that a TP of 10,000 less related individuals is necessary to achieve an accuracy of 0.7 for a trait of heritability 0.3. While Bernardo et al., 2007 reported that genomic selection outperformed both phenotypic and MARS when trait heritability was 0.2, 0.5 and 0.8, but GEBV accuracy decreased with decreasing heritability. Hence, it was considered that the range of heritability expressed here was adequate to test different scenarios as the ultimate scope of this study was to determine feasibility of GS in durum wheat.

Crossa et al., 2014 reported that not only heritability that affect the accuracy of the prediction, but also, the relationship between the individuals to be predicted and those used to train the models for prediction, number of markers, sample size and genotype environment interaction (GE).

IV-4-1. Selection of the best fitting statistical models for genomic predictions (i, ii, iii, iv)

The prediction analysis was conducted on the RILs population using models that account for the relationship matrix (G), environment effect (E), genotype by environment interaction (GxE), markers (M), and marker by environment interaction (MxE). The accuracy of breeding selection using only phenotypic data was computed (Figure IV-2) as G+E and GxE models (*i* and *ii*), to confirm that accuracies of 0.47-0.28 could be obtained via traditional breeding selection for GY. These results confirm what reported by Crossa et al. (2014) that pedigree (population structure) accounts for a sizeable proportion of the prediction accuracy. These values were set as competitors to determine the success of replacing phenotypic selection with molecular selection. Interestingly, the GS models that incorporated marker effect (*iii, iv*) generated non-significantly different or superior accuracies than traditional breeding selection, indicating a strong role for GS in future breeding (Figure IV-2).

IV-4-2. Size and relatedness of the training population (v, vi)

Beside academical studies, breeders often have limited resources and tend to reduce costs whenever possible. A decrease in the size of the TP that needs to be both

genotyped and phenotyped, and in the number of markers to be used for genotyping can represent important savings (Hinffer et al., 2011a, Crossa et al., 2014, Bassi et al., 2016). This possibility was tested by varying the proportion of individuals included in TP and VP from 75% TP and 25% VP, which is a very conservative and costly approach, to 50% TP and 50% VP, and even 25% TP and 75% VP. Interestingly, non-significant differences in accuracies could be observed for any of the reductions, for both high and low heritability traits (GY and TKW).

The relatedness between the TP and VP has been identified as a key consideration for predicting complex trait with low heritability. In an ideal scenario, breeders would like to accumulate information for a TP over time, using their normal yield trials as the source for this activity. By logic, the relatedness between such a TP and a BP under selection should be that of half-sibs. To test the feasibility of this approach, the four RIL populations that share half sib relationships were used to predict each other (Table IV-2). This resulted in severe losses of accuracy, reaching values close to zero for both high and low heritable traits (GY and TKW). This is in agreement with Windhausen et al. (2012) that also encountered accuracies close to zero when predicting far-related populations. The relatedness of a TP to the population to be predicted is hence one of the most critical aspect of GS in durum wheat. Therefore, small TP can be effectively deployed to accurately select BP only if these have full sibs relationships with the population to be selected. This is in good agreement with Bassi et al. (2016) that described several breeding schemes to deploy GS in a manner that would allow the TP to be full-sib of the BP under selection, without excessive loss of genetic gain.

IV-4-3. Does markers number affect the predictions? (v)

The possibility of deploying GS in breeding is still heavily hindered by the cost associated with genotyping huge populations. A way to reduce the cost of genotyping would be to reduce the number of markers used for the analysis. Here we tested the effect of the markers number to reveal that there was no significant difference in the prediction accuracies between using 3,202 or 320 SNPs as far as the TP and VP are full sibs (Figure IV-4). Hickey et al. (2014) also reported that when using information from related maize bi-parental populations high accuracies can be achieved using a small number of markers. Similarly, Haile et al. (2018) indicated that among advanced durum wheat breeding lines, the reduction from 9,000 to 500 markers did not cause a significant reduction in accuracies. However, it has to be noted that combining a decrease of TP size to 20% of the BP, and 10% of markers number caused the accuracy for GY to drop from 0.48 to 0.41 and for TKW from 0.77 to 0.74. This is a

significant reduction of 0.07 and 0.03 points. Still, in the optic of practical application, the values of accuracies remain very close to what achieved using only phenotypic models (G+E and GxE) and hence it could be advisable to deploy small TP and small markers set in breeding if this makes GS a more affordable approach.

IV-4-4. Is there an advantage to conduct QTL analysis before genomic predictions? (vii, viii)

QTL analysis and GS models rely on the same type of dataset. Therefore, it is of interest to define if there is additive contribution in combining both type of studies. Initially it was tested the effect of using only markers underlying QTLs (presented in annex 3) as predictor, as it would be the case when simulating a MAS approach (Figure IV-5). The obtained accuracies reached between -0.02 and 0.54, depending on traits and populations. This would suggest that running prediction models using only few markers linked to known genes (44 and 27 for GY and TKW, respectively) could provide some degree of success. For confirmation, the opposite situation was also tested by removing any markers associated to QTL from the whole dataset. Once again, the accuracies dropped significantly for all traits and populations, except SW. This result suggests that marker number is not the only factor to ensure high accuracies, but also the ability to define the haplotype of major effect loci is of critical importance.

The final test was designed to combine the extra information obtained via the definition of major allele effects by QTL analysis with the minor allele effects assessed via GS. Since the initial QTL discovery was conducted using the whole population, while GS models would instead use only sub-set of each population as TP and VP, QTL discovery was re-conducted for each TP subset. All initially identified QTLs were re-identified in 10-50% of the TP subsets (Annex 4) depending on the levels of allelic and phenotypic variation of each random subset. The marker underlying the QTL re-identified were fixed for each TP subset and used to improve the prediction model. The results are extremely promising since for all populations the combination of minor allele effect as GS random factor, and major allele effects as QTL fixed factor resulted in a significant increase in prediction accuracies. Furthermore, the accuracies value were increase by 0.06 to 0.12 points, a major increase compared to the 0.02 points of reducing the TP size or changing statistical models. Our results are in partial agreement with Sarinelli et al. (2019) that demonstrated that major genes added as fixed effects always improved model predictive ability, with the greatest gains coming from combinations of multiple genes for days to heading and plant height in a winter

wheat panel. Bian and Holland (2017) also concluded that adding SNP associated with a given trait as fixed effects resulted in higher predictive abilities when compared with models that only treat SNP as random effects. Bernardo (2014) pointed out that the prediction accuracy of GS models can be increased by adding major genes as fixed effects when they represent a large proportion of the total variance associated with the trait under consideration (≥ 10%). Considering that GY remains often the main targeted trait and also one of the most complex to predict, our results overall support the principle of incorporating fixed effect alleles into prediction model, especially for markers accounting for large part of the phenotypic variation. For practical breeding application, becomes then interesting the idea of combining MAS using marker associated to known loci as fixed effects, and all other loci as random effect. Furthermore, there appears to be additive value in conducting a discovery step via QTL analysis before running genomic predictions, since the additional information can be strategically exploited to increase accuracies.

IV-5. Conclusion

The results of this study provide a framework for better understanding and deploying molecular selection in durum wheat. The use of significant QTL for agronomic traits and incorporating them into prediction models to reveal significant gains of accuracy for GY when integrated as fixed effect. Several critical considerations were also tested for their deployment in durum wheat breeding. The results presented here are in good agreement with previous literature and what suggested previously by us for breeding application of GS in wheat (Bassi et al. 2016). In practice, the use of half sibs or distantly related TP does not appear to be an exploitable methodology for GS in durum wheat. Instead, small size full sibs TP needs to be deployed and genotyping costs can be reduced by using just 200-300 SNPs. In addition, known loci linked to traits of interest should be also included in the marker set and used as fixed effect to increase prediction. Most importantly, all genomic prediction models were compared to the accuracy attainable by classical phenotypic selection to confirm that the same results could be achieved via molecular approaches. Altogether, our result provides strong support for the deployment of genomic prediction in durum wheat breeding.

General discussion

Durum wheat (*Tritcum durum* Desf.) covered approximately 20 million hectares worldwide and accounts for around 6% of total wheat production (37.7 million tonnes in 2013; International Grain Council October 2014). Durum wheat is a staple crop in North Africa and West Asia (Elias and Manthey, 2005) and is becoming a strategic crop in sub-Sahara Africa with changes in food habits and urbanization of populations. Heat is the most important abiotic stresses affecting 58% of global wheat production area (Kosina et al., 2007). Temperatures are expected to raise in the near future due to climate change (IPCC, 2007; Ortiz et al., 2008). Coping with heat, requires the development of tolerant cultivars to maintain high yields.

Genetic diversity is of paramount importance as a source of novel traits and alleles for plant breeding, particularly to face the unpredictable challenges laying ahead, at a time of changing climates and new end-user demands (Tester et al., 2010). However, diversity *per se* is of limited use (Frankel et al., 1995: Royo et al., 2009). It is instead to the breeders' advantage to know which ideal sources of diversity should be integrated within each program. With this scope, the global diversity of durum wheat was assessed comparing breeding efforts, historical cultivars, and landraces from 28 countries. However, several factors can influence the genetic diversity within a germplasm collection and the analysis presented here can only explain a fraction of it. The results of the AMOVA confirmed that the DAPC model was able to capture approximately one fifth of the total variance by stratifying the panel into 10 clusters, with individuals that maintained high levels of genetic diversity within groups. Thus, even if the choice of k = 10 was conservative as shown by the AMOVA, it was considered adequate to better identify similarities between genotypes, rather than over-fit their differences (Jombart et al., 2010). As described above, the population stratification of this panel separates four groups of landraces and six groups of cultivars and elite lines. Cluster 1 includes landraces from Jordan, Syria and Iraq (Figure II-5), countries that correspond to the center of origin of durum wheat in the Levant (Mackey, 2005). The geographical proximity of these landraces to the center of origin maintained a high level of genetic purity with low levels of admixture (Figure II-5) and almost completes fixing of major (77%) and minor (10%) alleles (Table II-4). This is in good agreement with what was reported previously for landraces from Jordan (Rawashdeh et al., 2007; Mohammadi et al., 2014).

Ethiopia is known as a "secondary center of durum wheat diversity" (Harlan, 1969; Vavilov, 1992). Landraces from this country have unique morphology (Sakamoto and Fukui, 1972; Porceddu et al., 1973; Pecetti et al., 1992) and represent a separate sub-species under the name *T. durum* subs. abyssinicum or T. aethiopicum (Dejene et al., 2015; 2016). Figure II-5 clearly shows that this germplasm is distinct from the primary region of origin of durum wheat (Middle-East landraces) with substantially no kinship to it. Furthermore, there is limited admixture between this group and any others. Hence, Ethiopia truly represents a different center of diversity for durum wheat, without an evident allelic similarity to the primary origin in the Levantine, as also suggested by several other authors (Sakamoto and Fukui, 1972; Porceddu et al., 1973; Pecetti et al., 1992). Domesticated emmer reached Ethiopia some 5,000 years ago (National Research Council, 1996) probably arriving from Egypt along the Silk Road (Luo et al., 2007) and it occupies approximately 7% of the wheat production today under the name of aja. Thus, it can be suggested that Ethiopia is indeed not a secondary center of diversity, but rather a "secondary center of origin", where emmer was further domesticated to durum wheat as it occurred in the Levantine more than 7,000 years before.

In previous research, a panel of 190 Spanish durum wheat landraces was attributed to nine sub-populations (Ruiz et al., 2012), while a similar set of the germplasm collection used here, comprising 134 modern durum cultivars, was assigned to six sub-populations by Maccaferri et al. (2003). In our research, the number of clusters used for stratification could have been increased, to allocate four additional sub-populations among landraces of Cluster 4. However, the setting of k is highly dependent on the scope of the research and here the preference was given to capturing similarities rather than divergences. A large portion of admixture among landraces remained unfixed with the set value of *k*, and this could justify the difference in the number of clusters between our work and that of Ruiz et al. (2012). Similarly, the division into six clusters of modern material appeared in line with the results of previous authors (Maccaferri et al., 2003) and it provided interesting information about the history of alleles exchange among breeding programs. Slight differences were, however, observed from past works, due to the significant increase in this study of the number of elites derived from the ICARDA breeding program, which alone defined two novel well distinct clusters (5 and 9), and also the study of recent Australian and Canadian cultivars, which also created a cluster not described before by Maccaferri et al. (2003).

The collection described above, was then field-tested using north-south heat gradient. Where landraces couldn't face the extreme heat along the Senegal river. Thus, to conduct GWAS study, only 216 durum wheat modern lines from nine countries of origin including ICARDA and CIMMT lines were kept. There was a 24º-latitude difference between Moroccan stations and Senegal River stations. While all four stations experienced a significant heat stress throughout the growing cycle, the heat stress was less intense in Tessaout and Melk Zehr with maximum temperature around at flowering 26.6 and 30.4 °C during the second crop seasons, respectively. Fanaye and Kaedi were much warmer (maximum temperature: 37.1 and 39 °C during the second crop season, respectively) (Annex 2).

Agronomic performances for GY, TKW, and Gr.spk were recorded to reveal that Fanaye's mean yield was reduced by 35% and 52% compared to Melk Zhar during 2015 and 2016, respectively, and by 60% compared to Tessaout in 2016. While in Kaedi the mean yield was reduced by 52% and 46% compared to Melk Zhar (during 2015 and 2016, respectively) and by 55% compared to Tessaout in 2016. Similar results were reported previously by Sukumaran et al. (2018), who indicated that mean yield for CIMMYT durum wheat lines was reduced by 72% when the genotypes were grown under heat stress conditions.

The most surprising result was the excellent top yield achieved along the Senegal River, with productivities of up to 6.6 t ha^{-1} in Fanaye and 4.7 t ha^{-1} in Kaedi with seasons of just 92 days. In fact, the large kernels obtained along the River (47g in Kaedi and 49 g in Fanaye) allowed reaching high productivity even under a severe temperature stress. Also, Kumar et al. 2013 indicated that thousand kernel weight (TKW) has very important contributing against heat stress. However, Sall et al. (2018, 2019) failed to reveal this trait as critical for heat tolerance and rather indicated that grain setting (Gr.spk) was the most important.

heat susceptibility index (HSI) was estimated to compare between normal and stressed sites. Heat tolerance is a trait quantitative in nature, controlled by a large number of genes, which may be involved in interactions with each other (QTL × QTL interaction) and/or with the environments (QTL × E and QTL × QTL × E interactions) (Kumar et al., 2013). To dissect theses interactions, three environmental clusters were identified based on weather and phenological data, and used to estimate HIS corrected for GxE effects. While most of the studies that have been reported thus far used HSI under late-sown heat-stressed conditions and controlled conditions in soybean and wheat

(Ayeneh et al., 2002; Githiri et al., 2006; Kirigwi et al., 2007; Mohammadi et al., 2008; Pinto et al., 2010; Yang et al., 2010; Barakat et al., 2011; Mason et al., 2010, 2011), here actual heat-stressed environment was used to derive the contrasting value. Therefore, it is suggested by Mason et al. (2010) that HSI may be used as an indicator of yield stability and a proxy for heat tolerance.

Association mapping reveled seven major multiple traits QTLs on chromosomes 1A,2A, 4A, 4B, 5A, 5B and 7A with LODs ranged from 3.4 and 7.6 and explained phenotypic variation between 8 and 18%. They were expressed in the three clusters, regardless of heat stress pressure. Q.icd.DW.heat.14 was considered as heat-related QTLs and was identified in chromosome 4B and linked to the control of GY and the HSI of GY, TKW and Gr.m^{-2}. This result agrees with the finding of Sukumaran et al. (2018) that also identified a genomic region associated to HSI-GY in the same chromosome under heat pressure. The co-location of QTL for different agronomic and physiological traits with QTL for yield suggests that it is possible to achieve genetic dissection of the crop performance under heat stress to facilitate a more efficient breeding approach (Pinto et al., 2010). On chromosome 7B, a highly significant QTL ("Q.icd.DW.heat.23", LOD=4.0 and r^2=12%), was detected and covered the association to GY and the HSI of GY and TKW. previously, Paliwal et al.,2012, presented a similar QTL effect that control the HSI of GY and TKW on long arm of the same chromosome in hexaploid wheat under late sown conditions. Also within the same genomic region, Sukumaran et al. (2018) identified QTL that control tolerance index of the TKW on durum wheat CIMMYT lines under heat stress. This region should be kept in high consideration and go for marker validation. Furthermore, a KASP marker has already been validated to help accelerate the selection for the most important QTL identified here (Q.icd.Heat.11). with significant accuracy showed, it was possible to use it to select for the positive allele in a validation set of completely independent germplasm with different genetic backgrounds. As such, we believe its exploitation in breeding can begin with confidence until a better marker can be develop. Afterworld, Genomic predicting was estimated in four RILs populations of durum wheat, planted in four environments during the season 2014/2015. The total of 546 individuals were genotyped with 'genotyping by sequencing' (GBS) method, a consensus map was constructed and imputation was performed to generate a total of 3,202 polymorphic SNPs markers to conduct this study. Poland et al., 2012 indicated that no significant differences were observed in the accuracy of genomic-estimated

breeding values (GEBVs) among imputation methods, where the values of prediction accuracies with GBS were 0.28 to 0.45 for grain yield. Hence, the use of imputation method in this study should not have an effect on the overall quality of the analysis. The Bayesian ridge regression statistical model was used to carry out the genomic prediction for two complex traits: GY and TKW. This model approach is more appropriate than shrinkage regression in which there are few or no large effects and many small effects (Breiman, 1995), as is the case with most quantitative traits.
Crossa et al., 2014 reported that not only heritability that affect the accuracy of the prediction, but also, the relationship between the individuals to be predicted and those used to train the models for prediction, number of markers, sample size and genotype environment interaction (GE). Our results confirm what reported by Crossa et al. (2014) that pedigree (population structure) accounts for a sizeable proportion of the prediction accuracy.

breeders often have limited resources and tend to reduce costs whenever possible. A decrease in the size of the TP that needs to be both genotyped and phenotyped, and in the number of markers to be used for genotyping can represent important savings (Hinffer et al., 2011a, Crossa et al., 2014, Bassi et al., 2016). Interestingly, non-significant differences in accuracies could be observed for any of the reductions, for both high and low heritability traits (GY and TKW). To test the feasibility of the relatedness between TP and VP, the four RIL populations that share half sib relationships were used to predict each other (Table IV-2). This resulted in severe losses of accuracy, reaching values close to zero for both high and low heritable traits (GY and TKW). This is in agreement with Windhausen et al. (2012) that also encountered accuracies close to zero when predicting far-related populations. Therefore, small TP can be effectively deployed to accurately select BP only if these have full sibs relationships with the population to be selected. This is in good agreement with Bassi et al. (2016) that described several breeding schemes to deploy GS in a manner that would allow the TP to be full-sib of the BP under selection, without excessive loss of genetic gain.

we tested as well, the effect of the markers number to reveal that there was no significant difference in the prediction accuracies between using 3,202 or 320 SNPs as far as the TP and VP are full sibs (Figure IV-4). Hickey et al. (2014) also reported that when using information from related maize bi-parental populations high accuracies

can be achieved using a small number of markers. Similarly, Haile et al. (2018) indicated that among advanced durum wheat breeding lines.

QTL analysis and GS models rely on the same type of dataset. Therefore, it is of interest to define if there is additive contribution in combining both type of studies. Initially it was tested the effect of using only markers underlying QTLs (presented in annex 3) as predictor, as it would be the case when simulating a MAS approach (Figure IV-5). The obtained accuracies reached between -0.02 and 0.54, depending on traits and populations. This would suggest that running prediction models using only few markers linked to known genes (44 and 27 for GY and TKW, respectively) could provide some degree of success. Then, the marker underlying the QTL re-identified were fixed for each TP subset and used to improve the prediction model. The results are extremely promising since for all populations the combination of minor allele effect as GS random factor, and major allele effects as QTL fixed factor resulted in a significant increase in prediction accuracies. Our results are in partial agreement with Sarinelli et al. (2019) that demonstrated that major genes added as fixed effects always improved model predictive ability, with the greatest gains coming from combinations of multiple genes for days to heading and plant height in a winter wheat panel. Bernardo (2014) pointed out that the prediction accuracy of GS models can be increased by adding major genes as fixed effects when they represent a large proportion of the total variance associated with the trait under consideration (≥ 10%).Hence. There appears to be additive value in conducting a discovery step via QTL analysis before running genomic predictions, since the additional information can be strategically exploited to increase accuracies.

Conclusion and perspectives

The genetic diversity study (chapter II) showed that most of the modern germplasm has only 58 to 44 % of the genes still segregating, regardless of the breeding strategy or combination of germplasm utilized. Unexpectedly, the two centers of origin of durum wheat do not appear to be the most exploitable source of allelic diversity with most of their loci in fixed state. Rather, the landraces from Central and South Asia would provide the highest pool to integrate new alleles and should therefore be kept in high consideration for increasing diversity of modern breeding programs. Alternatively, the five clusters of ICARDA, CIMMYT, developed countries, 'Om Rabi' derivatives, and Italian breeding showed limited admixture with each other and therefore their inter-hybridization is a possible source of genetic diversity. Still, this will be possible only if the exchange of seeds for breeding purposes is kept free and unobstructed.

We have only 30 years left to completely revolutionize wheat cultivation, or the devastating effects of climate change will cause a drastic reduction in global supply. To accelerate the rate of genetic gain and delivery of superior varieties, the novel allele available in the ancient germplasm need to be coupled with methodology to rapidly integrate them into new cultivars. Molecular technologies such as genome wide association studies and genomic selection promise to be the turning point to give breeders a chance.

Instead of conventional direct selection for yield, genetic loci for yield components and associated traits can be identified to enable the rapid pyramiding of useful alleles. The MTAs identified in this study (Chapter III) are expected to benefit durum wheat breeders in selection of heat-tolerant plants. These can be used firstly to guide the crossing program. In fact, the normal breeding method relies on looking at the phenotypic performances of lines in the field and crossing the best to each other. However, it is possible that these equally performing lines have the same set of positive alleles and hence the crossing of the two would not allow any genetic gain. Instead, via GWAS it is now possible to determine which lines, once combined, would result in progenies with the highest possible number of positive alleles, which in turn would maximize rapid genetic gain. Secondly, GWAS can unlock the use of MAS. The same MTAs can then be converted into easy to use markers and deployed to further accelerate their pyramiding. Finally, the panel used for GWAS is both genotyped and

phenotyped, and it includes several key parental lines to be used for generating new breeding crosses. As such, it represents an ideal TP to be deployed via GS.
Genomic selection holds the potential to substantially accelerate the breeding cycle, enhancing gains per unit time. The study conducted here (Chapter IV) provides very valuable insight on how to conduct GS in durum wheat. The high accuracies obtained promote the rapid deployment of this technology into practical breeding. At the same time, the clear inability of non-full-sibs TP to predict performances imposes to breeders a more conservative use by designing specific schemes to reach rapid gain.

Together, the work of this Thesis was conducted across five countries (Morocco, Lebanon, Senegal, Mauritania and Sweden), and it generated both novel scientific knowledge for advancing the field as well as extremely practical information to be immediately converted into breeding advancements. Hence, it provides a small step forward in the fight against climate change.

Still, given the time lag between technology development, deployment, and on-farm adoption, this research will not yield its positive effects if institutional innovations and policy options are not put in place to facilitate farmers' access to existing and new germplasm

References

- Acosta-Gallegos, J.A., Kelly, J.D. and Gepts, P. (2007) Pre-breeding and genetic diversity in common bean (Phaseolus vulgaris). *Crop Sci.*, 47, S44–S59.
- Agricultural and Resource Economics, Commonwealth of Australia: Canberra,
- Åkerberg, E. (1986). *Nilsson-Ehle and the development of plant breeding at Svalöf during the period 1900–1915*. Hereditas, 105(1), 1-5.
- Altintas, S., Toklu, F., Kafkas, S., Kilian, B., Brandolini, A. and Ozkan, H. (2008) Estimating genetic diversity in durum and bread wheat cultivars from Turkey using AFLP and SAMPL markers. *Plant Breed.,* 127, 9–14.
- Alqudah, A. M., Sallam, A., Baenziger, P. S., & Börner, A. (2020). GWAS: Fast-forwarding gene identification and characterization in temperate Cereals: lessons from Barley–A review. *Journal of Advanced Research,* 22, 119.
- Anon. (2009) 'Australian crop report 16 June 2009 no. 150.' (Australian Bureau of
- Anonymous. 2006. La biologie de Triticum turgidum ssp. Durum (Blé dur).
- Aseng, J. V., Rantung, J., & Poeng, R. (2015). Optimasi biaya proses karena variasi kecepatan dan kedalaman potong pada poros aisi-1040 menggunakan mesin bubut knuth dm-1000a. *jurnal online poros teknik mesin unsrat,* 4(1).
- Ashraf, M., Akram, N. A., Arteca, R. N., & Foolad, M. R. (2010). The physiological, biochemical and molecular roles of brassinosteroids and salicylic acid in plant processes and salt tolerance. *Critical Reviews in Plant Sciences,* 29(3), 162-190.
- Asima, G., Dar, Z.A., Lone, A.A. Yasin, A., Ali, Y. and Habib, M. (2017). Breeding climate change resilient maize and wheat for food security. *Archives of Agriculture and Environmental Science*, 2(2): 129-133.
- Asima, Gazal, et al. (2017). "Breeding climate change resilient maize and wheat for food security." *Archives of Agriculture and Environmental Science 2.2 129-133.*
- Asoro, F. G., Newell, M. A., Beavis, W. D., Scott, M. P., & Jannink, J. L. (2011). Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *The Plant Genome,* 4(2), 132-144.
- Autrique, E., Nachit, M.M., Monneveux, P., Tanksley, S.D. and Sorrells, M.E. (1996) Genetic diversity in durum wheat based on RFLPs, morphophysiological traits, and coefficient of parentage. *Crop Sci.*, 36, 735–742.
- Ayeneh, A., Van Ginkel, M., Reynolds, M. P., & Ammar, K. (2002). Comparison of leaf, spike, peduncle and canopy temperature depression in wheat under heat stress. *Field Crops Research*, 79(2-3), 173-184.
- Baldy C. 1984. Agro météorologie des cultures multiples en régions chaudes. *Ed INRAF*, 120 pp.
- Balla K, Bencze S, Bónis P, Árendás T, Veisz O (2014). Changes in the photosynthetic efficiency of winter wheat in response to abiotic stress. *Cent. Eur. J. Biol.*, 9: 519-530.
- Barakat Najeb, M., Al-Doss, A. A., Elshafei, A. A., & Moustafa, K. A. (2011). Identification of new microsatellite marker linked to the grain filling rate as indicator for heat tolerance genes in F2 wheat population. *Australian Journal of crop science*, 5(2), 104.
- Bari, A., Street, K., Mackay, M., Endresen, D.T.F., De Pauw, E. and Ahmed, A. (2012) Focused identification of germplasm strategy (FIGS) detects wheat stem rust resistance linked to environmental variables. *Genet. Resour. Crop Evol.*, 59, 1465 1481.
- Barlow KM, Christy BP, O'Leary GJ, Riffkin PA, Nuttall JG. 2015. Simulating the impact of extreme heat and frost events on wheat crop production: a review. *Field Crops Research* 171: 109-119.
- Barnabas B, Jager K, Feher A. 2008. The effect of drought and heat stress on reproductive processes in cereals. *Plant Cell Environ* 31: 11-38.
- Bassi FM, Bentley AR, Charmet G, Ortiz R, et al. (2016). Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). *Plant Sci.* 242: 23-36
- Battisti, David S., and Rosamond L. Naylor. (2009) "Historical warnings of future food insecurity with unprecedented seasonal heat." *Science,* 323.5911: 240-244.

- Beavis, W. D. (1994, December). The power and deceit of QTL experiments: lessons from comparative QTL studies. In Proceedings of the forty-ninth annual corn and sorghum *industry research conference* (pp. 250-266).
- Bernardo, R., & Charcosset, A. (2006). Usefulness of gene information in marker-assisted recurrent selection: a simulation appraisal. *Crop Science*, 46(2), 614-621.
- Bhusal, N., Sarial, A. K., Sharma, P., & Sareen, S. (2017). Mapping QTLs for grain yield components in wheat under heat stress. *PloS one*, 12(12), e0189594.
- Blanco, A., Bellomo, M. P., Cenci, A., De Giovanni, C., D'ovidio, R., Iacono, E., ... & Simeone, R. (1998). A genetic linkage map of durum wheat. *Theoretical and Applied Genetics*, 97(5-6), 721-728.
- Bohnert HJ, Gong Q, Li P, Ma S (2006) Unraveling abiotic stress tolerance mechanisms—getting genomics going. *Curr Opin Plant Biol* ,9:180–188
- Bonjean, A.P., W. J. Angus, M. van Ginkel. (2016) *The world wheat book: a history of wheat breeding* Volume 3. Lavoisier.
- Botstein, D. and Risch, N. 2003. Discovering genotypes underlying human phenotypes: Past successes for Mendelian disease, future approaches for complex disease. *Nat. Genet.* 33 Suppl: 228–237
- Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.*, 32, 314-331.
- Bozzini A. 1988. "Origin, distribution, and production of durum wheat in the world". Dans Fabriani G. et C. Lintas (éd). Durum: Chemistry and Technology. *AACC (Minnesota)*, États-Unis. p. 1-16.
- Bozzini, A. (1988) Origin, distribution, and production of durum wheat in the world. In: Durum: chemistry and technology (Dans Fabriani, G. and Lintas, C., eds.), pp. 1–16. *St. Paul, Minnesota*: American Association of Cereal Chemists.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics,* 23(19), 2633-2635.
- Bradshaw, J.E. and Ramsay, G. (2005) Utilization of the commonwealth potato collection in potato breeding. *Euphytica* ,146, 9–19.
- Burguen˜ o, J. et al. (2012) Genomic prediction of breeding values when modeling genotype _ environment interaction using pedigree and dense molecular markers. *Crop Sci.* 52, 707–719
- Burgueño, J., de los Campos, G., Weigel, K., & Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype× environment interaction using pedigree and dense molecular markers. *Crop Science*, 52(2), 707-719.
- Cabrera, A., Souza, E., Guttieri, M., Sturbaum, A., Hoffstetter, A. and Sneller, C. (2015) Genetic diversity, linkage disequilibrium and genome evolution in soft winter wheat. *Crop Sci.*, 54, 2433–2448.
- Cantrell, R.G., Janick, J. and Nostrand Reinhol Co (1987) Breeding and genetics of durum wheat. *Plant Breeding Reviews*, New York, 5, 11-40.
- Chakraborty, R., & Weiss, K. M. (1988). Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proceedings of the National Academy of Sciences*, 85(23), 9119-9123.
- Charmet, G., Robert, N., Perretant, M. R., Gay, G., Sourdille, P., Groos, C., ... & Bernard, M. (1999). Marker-assisted recurrent selection for cumulating additive and interactive QTLs in recombinant inbred lines. *Theoretical and Applied Genetics*, 99(7-8), 1143-1148.
- Christiansen, M.J., Andersen, S.B. and Rodomiro, O. (2002) Diversity changes in an intensively bred wheat germplasm during the 20th century. *Mol. Breed.*, 9, 1–11. CIMMYT. Mexico D.F., Mexico.
- Combe L, Picard D. 1994. Elaboration du rendement des principales cultures annuelles. Paris INRA.
- Cooper, H.D., Spillane, C., Hodgkin, T., eds. (2001) Broadening the genetic base of crop production. Wallingford, Oxon, United Kingdom: *CABI Publishing*.

- Crosbie, T.M., S.R. Eathington, G.R. Johnson, M. Edwards, R.Reiter, S. Stark, et al. 2003. Plant breeding: Past, present, and future. p. 1–50. In K.R. Lamkey and M. Lee (ed.) *Plant Breeding*: The Arnel R. Hallauer Int. Symp., Mexico City.17–23 Aug. 2003. Blackwell, Oxford, UK.
- Crossa, J., de Los Campos, G., Pérez, P., Gianola, D., Burgueño, J., Araus, J. L., ... & Arief, V. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, 186(2), 713-724.
- Crossa, J., P. Peirez, J. Hickey, and J. BurguenTo. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity (Edinb)* 112:48–60. doi:10.1038/hdy.2013.16
- Crossa, J., Pérez, P., de los Campos, G., Mahuku, G., Dreisigacker, S., & Magorokosho, C. (2011). Genomic selection and prediction in plant breeding. *Journal of Crop Improvement*, 25(3), 239-261.
- Crossa, J., Perez, P., Hickey, J., Burgueno, J., Ornella, L., Cerón-Rojas, J., ... & Bonnett, D. (2014). Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity*, 112(1), 48.
- Daetwyler, H. D., Pong-Wong, R., Villanueva, B., & Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics*, 185(3), 1021-1031.
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7), 499.
- De Los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., & Calus, M. P. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, 193(2), 327-345.
- De Los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., ... & Cotes, J. M. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*, 182(1), 375-385.
- Dekkers, J. C., & Hospital, F. (2002). Multifactorial genetics: The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics*, 3(1), 22.
- Delêtre, M., McKey, D.B. and Hodkinson, T.R. (2011) Marriage exchanges, seed exchanges, and the dynamics of manioc diversity. *Proc. Natl. Acad. Sci.* USA,108, 18249–18254. Dept Agriculture, Washington, DC No. 180:1–63.
- Deu, M., Sagnard, F., Chantereau, J., Catalayud, C., Hérault, D., Mariac, C. and Pham, J.L. (2013) Diversification of primary gene pool and genetic enhancement of grain legumes. *Front. Plant Sci.*, 4, 309.
- Dias AS, Lidon FC. 2009. Evaluation of grain filling rate and duration in bread and durum wheat under heat stress after anthesis. *Journal of Agronomy and Crop Science* 195:137-147.
- Dias, S., N. J. Welton, D. M. Caldwell, and A. E. Ades. "Checking consistency in mixed treatment comparison meta-analysis." Statistics in medicine 29, no. 7-8 (2010): 932-944.
- Driedonks, N., Rieu, I., & Vriezen, W. H. (2016). Breeding for plant heat tolerance at vegetative and reproductive stages. *Plant reproduction*, 29(1-2), 67-79.
- Dyer, G.A. and Taylor, J.E. (2008) A crop population perspective on maize seed system in Mexico. *Proc. Natl. Acad. Sci.* USA, 105, 470–475.
- Kirby E. J. M and Margaret Appleyard, Cereal Development Guide, 2d ed. (Warwickshire, England: *National Agricultural Centre*, 1987), pp. 4, 12, 13, 14.
- Kosina, P., Reynolds, M., Dixon, J., & Joshi, A. (2007). Stakeholder perception of wheat production constraints, capacity building needs, and research partnerships in developing countries. *Euphytica*, 157(3), 475-483.
- Easterling, W. E. (1996). Adapting North American agriculture to climate change in review. *Agricultural and Forest Meteorology*, 80(1), 1-53.
- Elias, E. M., Manthey, F. A., Stack, R. W., & Kianian, S. F. (2005, December). Breeding efforts to develop Fusarium head blight resistant durum wheat in North Dakota. *In Proc* (pp. 25-26).

- Excoffier, L. Laval, G. and Schneider, S. (2005). Arlequin ver. 3.5.2.2: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1, 47–50.
- Falush, D., Stephens, M. and Pritchard, J.K. (2003) Inference of population structure using multilocus genotype data, linked loci and correlated allele frequencies. *Genetics*, 164, 1567–1587.
- Feldman, M. (2001) "Origin of cultivated wheat," In: The World Wheat Book: A History of Wheat Breeding (Dans Bonjean, A.P, and Angus W.J., eds.), pp. 3–58. Andover, United Kingdom: Intercept Limited.
- Fisher, R. A. (1918). 009: The Correlation Between Relatives on the Supposition of Mendelian Inheritance.
- Fisher, R. A., & Maurer, R. (1978). Drought resistance in spring wheat cultivars. I. Grain yield responses in spring wheat. *Australian J. Agric. Sci*, 29, 892-912.
- Flint-Garcia, S.A., Thornsberry, J.M. and Buckler, E.S. (2003) Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.*, 54, 357–374.
- Foulkes MJ, Slafer GA, Davies WJ, Berry PM, Sylvester-Bradley R, Martre P, Calderini DF, Giffiths S, Reynolds MP (2011). Raising yield potential of wheat. (III) Optimizing partitioning to grain while maintaining lodging resistance. *Journal of Experimental Botany* 62,469–486
- Frankel, OH., Brown, AH., Burdon, JJ., (1995) The conservation of plant biodiversity: *Cambridge University Press*.
- Gale, M. D., and Youssefian, S. (1985). "Dwarfing genes in wheat," In: *Progress in Plant Breeding* (Russell, G. E., ed.), pp. 1–35. Butterworths, London.
- Gate PH. 1995. Ecophysiologie du blé ; Technique et documentation : *Lavoisier*, Paris 429 p.
- Gioia, T., Nagel, K.A., Beleggia, R., Fragasso, M., Ficco, D.B.M., Pieruschka, R., De Vita, P., Fiorani, F. and Papa, R. (2015) Impact of domestication on the phenotypic architecture of durum wheat under contrasting nitrogen fertilization. *J. Exp. Bot.*, 66, 5519–5530.
- Githiri, S. M., Watanabe, S., Harada, K., & Takahashi, R. (2006). QTL analysis of flooding tolerance in soybean at an early vegetative growth stage. *Plant Breeding*, 125(6), 613-618.
- Goddard, M. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica (The Hague)* 136:245–257
- Golabadi, M., Arzani, A. and Maibody, S.A.M. (2006) Assessment of drought tolerance in segregating populations in durum wheat. *Afr. J. Agric. Res.*, 5, 162–171.
- González-Camacho, J. M., de Los Campos, G., Pérez, P., Gianola, D., Cairns, J. E., Mahuku, G., ... & Crossa, J. (2012). Genome-enabled prediction of genetic values using radial basis function neural networks. *Theoretical and Applied Genetics*, 125(4), 759-771.
- Goulden, c. H. 1939. Problems in plant selection. In Proceedings of the Seventh International Genetics Congress. *Cambridge University Press*, pp. 132-133.
- Hagenblad, J., Zie, J. and Leino, M.W. (2011) Exploring the population genetics of gene bank and historical landrace varieties. *Genetic Resour. Crop Evol.*, 59, 1185–1199.
- Hamblin, M. T., Buckler, E. S., & Jannink, J. L. (2011). Population genetics of genomics-based crop improvement methods. *Trends in Genetics*, 27(3), 98-106.
- Hansen, J., Sato, M., & Ruedy, R. (2012). Perception of climate change. Proceedings of the National Academy of Sciences, 109(37), E2415-E2423.
- Harlan, J.R. (1969) Ethiopia: a center of diversity. *Econ. Bot.*, 23, 309–314.
- Hayes, B.J., P.J. Bowman, A.J. Chamberlain, and M.E. Goddard. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92:433–443.
- Heffner E, Jannink J, Iwata H, Souza E, Sorrells M (2011a) Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Science* 51(6):2597–2606
- Heffner E, Jannink J, Sorrells M (2011b) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *The Plant Genome* 4(1):65–75
- Heffner, E. L., Sorrells, M. E., & Jannink, J. L. (2009). Genomic selection for crop improvement. *Crop Science*, 49(1), 1-12.

- Hickey, J. M., Crossa, J., Babu, R., & de los Campos, G. (2012). Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. *Crop Science*, 52(2), 654-663.
- Hickey, J. M., Dreisigacker, S., Crossa, J., Hearne, S., Babu, R., Prasanna, B. M., ... & Gorjanc, G. (2014). Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Science*, 54(4), 1476-1488.
- Hijmans, Robert J., Luigi, G. and Mathur, P. (2012) DIVA-GIS version 7.5 manual.
- Hoisington, D., Khairallah, M., Reeves, T., Ribaut, J.M., Skovmand, B., Taba, S. and Warburton, M. (1999) Plant genetic resources: What can they contribute toward increased crop productivity? *Proc. Natl. Acad. Sci.* USA, 96, 5937–5943.
- Howarth C.J., 2005. Genetic improvements of tolerance to high temperature. In: Abiotic Stresses Plant Resistance Through Breeding and Molecular Approaches (Eds M. Ashraf, P.J.C. Harris). *Howarth Press*, New York, USA
- International Wheat Council (1991). World production durum wheat. The Council, London.
- IPCC. 2007. Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Core Writing Team, Pachauri RK, Reisinger A, eds. Geneva: Intergovernmental Panel on Climate Change. http://www.ipcc.ch/publications_and_data/publications_and_data_reports.shtml. Last accessed 2 January 2015.
- IPCC. 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM, eds. Cambridge, UK/New York: *Cambridge University Press*. http://www.ipcc.ch/publications_and_data/publications_and_data_reports.shtml Last accessed 2 January 2015.
- IWGSC, I. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome by the international wheat genome sequencing consortium (iwgsc). *Science*, 361.
- Jombart, T. and Collins, C. (2015). A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0. Imperial College London
- Jombart, T., Devillard, S. and Balloux, F. (2010) Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, 11, 94.
- Jonas, E., & de Koning, D. J. (2013). Does genomic selection have a future in plant breeding? *Trends in biotechnology*, 31(9), 497-504.
- Kabbaj, H., Sall, A. T., Al-Abdallat, A., Geleta, M., Amri, A., Filali-Maltouf, A., ... & Bassi, F. M. (2017). Genetic diversity within a global panel of durum wheat (Triticum durum) landraces and modern germplasm reveals the history of alleles exchange*. Frontiers in plant science*, 8, 1277.
- Karagöz, A. and Zencirci, N. (2005) Variation in wheat (Triticum spp.) landraces from different altitudes of three regions of Turkey. *Genet. Resour. Crop Evol.*, 52, 775–785.
- Kaur, V., & Behl, R. (2010). Grain yield in wheat as affected by short periods of high temperature, drought and their interaction during pre-and post-anthesis stages. *Cereal Research Communications*, 38(4), 514-520.
- Kidane, Y. G., Mancini, C., Mengistu, D. K., Frascaroli, E., Fadda, C., Pè, M. E., & Dell'Acqua, M. (2017). Genome wide association study to identify the genetic base of smallholder farmer preferences of durum wheat traits. Frontiers in plant science, 8, 1230.
- Kirigwi, F. M., Van Ginkel, M., Brown-Guedira, G., Gill, B. S., Paulsen, G. M., & Fritz, A. K. (2007). Markers associated with a QTL for grain yield in wheat under drought. *Molecular Breeding*, 20(4), 401-413.
- Knott, D.R. and J. Kumar. 1975. Comparison of early generation yield testing and single seed descent procedure in wheat breeding. *Crop Sci.*15: 295-299.
- Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature genetics*, 22(2), 139.

- Kumar, N., Markar, S., & Kumar, V. (2014). Studies on heritability and genetic advance estimates in timely sown bread wheat (Triticum aestivum L.). *Bioscience Discovery*, 5(1), 64-69.
- Kumar, R. R., Sharma, S. K., Goswami, S., Singh, G. P., Singh, R., Singh, K., ... & Rai, R. D. (2013). Characterization of differentially expressed stress-associated proteins in starch granule development under heat stress in wheat (Triticum aestivum L.).
- Kumari, K., Muthamilarasan, M., Misra, G., Gupta, S., Subramanian, A., Parida, S. K., ... & Prasad, M. (2013). Development of eSSR-markers in Setaria italica and their applicability in studying genetic diversity, cross-transferability and comparative mapping in millet and non-millet species. *PloS one*, 8(6), e67742.
- Laidò, G., Mangini, G., Taranto, F., Gadaleta, A., Blanco, A., Cattivelli, L., Marone, D., Mastrangelo, A.M., Papa, R. and De Vita, P. (2013) Genetic diversity and population structure of tetraploid wheats (Triticum turgidum L.) estimated by SSR, DArT and pedigree data. *PloS One*, 8:e67280.
- Lantican, M.A., Payne T.S., Sonder, K., Singh, R., van Ginkel, M., Baum, M., Braun, H.J., and Erenstein, O. (2015) Impacts of International Wheat Improvement Research in the World, 1994-2014. Mexico, D.F.: CIMMYT.
- LMC International (2009) Evaluation of measures relating to the durum wheat sector within the context of the common agricultural policy. NewYork, Oxford, Kuala Lumpur: *LMC International*.
- Lobell, D. B., Bänziger, M., Magorokosho, C., & Vivek, B. (2011). Nonlinear heat effects on African maize as evidenced by historical yield trials. *Nature climate change*, 1(1), 42.
- Lobell, David B., Uris Lantz C. Baldos, and Thomas W. Hertel. (2013) "Climate adaptation as mitigation: the case of agricultural investments." *Environmental Research Letters 8.1*: 015012.
- Lopez-Cruz, M., Crossa, J., Bonnett, D., Dreisigacker, S., Poland, J., Jannink, J. L., ... & de los Campos, G. (2015). Increased prediction accuracy in wheat breeding trials using a marker× environment interaction genomic selection model. G3: *Genes, Genomes, Genetics, 5(4), 569-582.*
- Lorenzana, R.E., and R. Bernardo. (2009). Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor. Appl. Genet.* 120:151–161. doi:10.1007/s00122-009-1166-3.
- Lotze-Campen, H., & Schellnhuber, H. J. (2009). Climate impacts and adaptation options in agriculture: what we know and what we don't know. *Journal für verbraucherschutz und Lebensmittelsicherheit*, 4(2), 145-150.
- Mac Key, J.M. (2005). "Wheat: its concept, evolution, and taxonomy," In: Durum Wheat Breeding: Current Approaches and Future Strategies Vol.1 (Royo, C., Nachit, M.M., Di Fonzo, N., Araus, J.L., Pfeiffer, W.H. and Slafer, G.A., eds.), pp. 3–61. New York, London, Oxford: Food Products Press, an imprint of the Haworth Press Inc.
- Maccaferri, M., Harris, N. S., Twardziok, S. O., Pasam, R. K., Gundlach, H., Spannagl, M., ... & Himmelbach, A. (2019). Durum wheat genome highlights past domestication signatures and future improvement targets. *Nature genetics*, 51(5), 885.
- Maccaferri, M., Sanguineti, M.C., Donini, P. and Tuberosa, R. (2003) Microsatellite analysis reveals a progressive widening of the genetic basis in the elite durum wheat germplasm. *Theor. Appl. Genet.*, 107, 783–797.
- Mackay, M., von Bothmer, R. and Skovmand, B. (2005) Conservation and utilization of plant genetic resources – future directions. *Czech J. Genet. Plant Breed.*, 41, 335–344.
- Maier, U. (1996) Morphological studies of free-threshing wheat ears from a Neolithic site in southwest Germany, and the history of the naked wheats. *Vegetation History Archaeobotany*, 5, 39–55.
- Martos, V., Royo, C., Rharrabti, Y. and Garcia del Morala, L.F. (2005) Using AFLPs to determine phylogenetic relationships and genetic erosion in durum wheat cultivars released in Italy and Spain throughout the 20th century. *Field Crop Res.*, 91, 107–116.
- Masle MJ. 1981. Relation entre croissance et développement pendant la montaison d'un peuplement de blé d'hiver, influence des conditions de nutrition. *Agronomie 1*: 365-374.

- Masle MJ. 1982. Mise en évidence d'un stade critique par la montée d'une talle. *Agronomie 1* : 623-632.
- Mason, R. E., Mondal, S., Beecher, F. W., & Hays, D. B. (2011). Genetic loci linking improved heat tolerance in wheat (Triticum aestivum L.) to lower leaf and spike temperatures under controlled conditions. *Euphytica*, 180(2), 181-194.
- Mason, S., McNeill, A., McLaughlin, M. J., & Zhang, H. (2010). Prediction of wheat response to an application of phosphorus under field conditions using diffusive gradients in thin-films (DGT) and extraction methods. *Plant and soil*, 337(1-2), 243-258.
- Melchinger, A. E., Utz, H. F., & Schön, C. C. (1998). Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics*, 149(1), 383-403.
- Mengistu, D.K. Kidane, Y.G., Catellani, M., Frascaroli, E., Fadda, C., Pe, M.E. and Dell'Acqua, M. (2016) High-density molecular characterization and association mapping in Ethiopian durum wheat landraces reveals high diversity and potential for wheat breeding. *Plant Biotechnol. J.*, 14, 1800–1812.
- Mengistu, D.K., Kiros, A.Y. and, Pè, M.E. (2015) Phenotypic diversity in ethiopian durum wheat (Triticum turgidum var. durum) landraces. *Crop J.*, 3, 190–199.
- Meuwissen, T. (2007). Genomic selection: marker assisted selection on a genome wide scale. *Journal of animal Breeding and genetics*, 124(6), 321-322.
- Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819.
- Mohammadi V, Qannadha MR, Zali AA, Yazdi-Samadi B. 2004. Effect of post anthesis heat stress on head traits of wheat. *Internat J Agric Biol* 6:42-44.
- Mohammadi, R., Haghparast, R., Sadeghzadeh, B., Ahmadi, H., Solimani, K. and Amri, A. (2014) Adaptation patterns and yield stability of durum wheat landraces to highland cold rainfed areas of Iran. *Crop Sci.*, 54, 944–954.
- Mohammadi, V., Zali, A. A., & Bihamta, M. R. (2008). Mapping QTLs for heat tolerance in wheat. *Mol. Breed.* 2011, 28, 281–301.
- Moragues, M., del Moral, L. F. G., Moralejo, M., & Royo, C. (2006). Yield formation strategies of durum wheat landraces with distinct pattern of dispersal within the Mediterranean basin I: Yield components. *Field Crops Research*, 95(2-3), 194-205.
- Moragues, M., Moralejo, M., Sorrells, M. E., & Royo, C. (2007). Dispersal of durum wheat [Triticum turgidum L. ssp. turgidum convar. durum (Desf.) MacKey] landraces across the Mediterranean basin assessed by AFLPs and microsatellites. *Genetic resources and crop evolution*, 54(5), 1133-1144.
- N'Diaye, A., Haile, J. K., Cory, A. T., Clarke, F. R., Clarke, J. M., Knox, R. E., & Pozniak, C. J. (2017). Single marker and haplotype-based association analysis of semolina and pasta colour in elite durum wheat breeding lines using a high-density consensus map. *PloS one*, 12(1), e0170941.
- Nachit, M., G., Ketata, H., Gauch, H. G., & Zobel, R. W. (1992). Use of AMMI and linear regression models to analyze genotype-environment interaction in durum wheat. *Theoretical and Applied genetics*, 83(5), 597-601.
- Nachit M.M. (1997) Network for integrated research to improve durum wheat productivity in the Mediterranean region. In: Braun HJ., Altay F., Kronstad W.E., Beniwal S.P.S., McNab A. (eds) Wheat: Prospects for Global Improvement. Developments in Plant Breeding, vol 6. Springer, Dordrecht. https://doi.org/10.1007/978-94-011-4896-2_73
- Nakaya, A., & Isobe, S. N. (2012). Will genomic selection be a practical method for plant breeding? *Annals of botany*, 110(6), 1303-1316.
- National Research Council (1996). Lost Crops of Africa volume I: grains. Washington, DC: The National Academies Press.
- Nesbitt, M. and Samuel, D. (1998) Wheat domestication: archaeobotanical evidence. Science, 279, 5356.1431e.
- Noohi K, Fatahi E, Kamali GHA. 2009. Heat stress effects analysis on wheat crop in southern provinces. Geophysical Res Abstracts 11:4441.

- Olesen, J. E., & Bindi, M. (2002). Consequences of climate change for European agricultural productivity, land use and policy. *European journal of agronomy*, 16(4), 239-262.
- Ortiz, R., Sayre, K. D., Govaerts, B., Gupta, R., Subbarao, G. V., Ban, T., ... & Reynolds, M. (2008). Climate change: can wheat beat the heat?. *Agriculture, Ecosystems & Environment*, 126(1-2), 46-58.
- Ortiz, R., Trethowan, R., Ortiz Ferrara, G., Iwanaga, M., Dodds ,J.H., Crouch, J.H., Crossa, J. and Braun, H.-J. (2007) High yield potential, shuttle breeding and a new international wheat improvement strategy. *Euphytica*, 157, 365–384.
- Ozkan, H., Levy, A. A., & Feldman, M. (2002). Rapid differentiation of homeologous chromosomes in newly-formed allopolyploid wheat. *Israel journal of plant sciences*, 50(sup1), 65-76.
- Paliwal, R., Röder, M. S., Kumar, U., Srivastava, J. P., & Joshi, A. K. (2012). QTL mapping of terminal heat tolerance in hexaploid wheat (T. aestivum L.). *Theoretical and Applied Genetics*, 125(3), 561-575.
- Parry MAJ, Reynolds M, Salvucci ME, Raines C, Andralojc PJ, Zhu X-G, Price GD, Condon AG, Furbank RT. 2011. Raising yield potential of wheat. II. Increasing photosynthetic capacity and efficiency. *Journal of Experimental Botany* 62, 453–467.
- Payne, R. W. (2009). GenStat. Wiley Interdisciplinary Reviews: Computational Statistics, 1(2), 255-258.
- Pecetti, L and Annicchiarico, P. (1998) Agronomic value and plant type of Italian durum wheat cultivars from different eras of breeding. *Euphytica*, 99, 9–15.
- Pecetti, L, Annicchiarico, P. and Damania, A.B. (1992) Biodiversity in a germplasm collection of durum wheat. *Euphytica*, 60, 229–238.
- Peng, J. H., Sun, D., & Nevo, E. (2011). Domestication evolution, genetics and genomics in wheat. *Molecular Breeding*, 28(3), 281.
- Pérez, P., & de los Campos, G. (2013). BGLR: a statistical package for whole genome regression and prediction. R package version, 1(0.2).
- Pérez-Rodríguez, P., Gianola, D., González-Camacho, J. M., Crossa, J., Manès, Y., & Dreisigacker, S. (2012). Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. *G3: Genes, Genomes, Genetics*, 2(12), 1595-1605.
- Perrier, X., Flori, A., Bonnot, F. (2003). “Data analysis methods,” In: Genetic diversity of cultivated tropical plants (Hamon, P., Seguin, M., Perrier, X., Glaszmann, J. C., eds), Enfield, pp 43 – 76. *Science Publishers*. Montpellier.
- Pearson, K. Notes on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.* 1985, 58, 240–242.
- Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2011). the R Development Core Team (2013) nlme: linear and nonlinear mixed effects models. R package version, 3, 1-97.
- Pinto, R. S., Reynolds, M. P., Mathews, K. L., McIntyre, C. L., Olivares-Villegas, J. J., & Chapman, S. C. (2010). Heat and drought adaptive QTL in a wheat population designed to minimize confounding agronomic effects. *Theoretical and Applied Genetics*, 121(6), 1001-1021.
- Porceddu, E., Perrino, P. and Olita, G. (1973). “Preliminary information on an Ethiopian wheat germplasm collection mission,” In: Proceedings of the Symposium on Genetic and Breeding of Durum Wheat (Scarascia Mugnozza, G.T., ed.), pp. 181–200. Bari, Italy: University of Bari.
- Porter, J. R., & Semenov, M. A. (2005). Crop responses to climatic variation. Philosophical Transactions of the Royal Society B: *Biological Sciences*, 360(1463), 2021-2035.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-959.
- Pritchard, J.K. and Rosenberg, N.A. (1999) Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum. Genet.*, 65, 220–228.

- Pusadee, T., Jamjod, S., Chiang, Y.-C., Rerkasem, B. and Schaal, B.A. (2009) Genetic structure and isolation by distance in a landrace of Thai rice. *Proc. Natl. Acad. Sci.* USA, 106, 13880–13885.
- R Development Core Team (2011) R: A Language and Environment for Statistical Computing. Vienna, Austria: The R Foundation for Statistical Computing.
- R Development Core Team. 2009. R: A language and environment for statistical computing [Online.] Available at: http://www.Rproject.org (verifi ed 18 June 2010). R Foundation for Statistical Computing. ISBN 3-900051-07-0. Vienna, Austria.
- Raman, H., Stodart, B.J., Cavanagh, C., Mackay, M., Morell, M., Milgate, A. and Peter, M. (2010) Molecular diversity and genetic structure of modern and traditional landrace cultivars of wheat (Triticum aestivum L.). *Crop Pasture Sci.*, 61, 222–229.
- Rane J, Pannu RK, Sohu VS, Saini RS, Mishra B, Shoran J, Crossa J, Vargas M, Joshi K (2007) Performance of yield and stability of advanced wheat cultivar under heat stress environments of the indo-gangetic plains. *Crop Sci.* 47:1561-1572.
- Rawashdeh, N.K., Haddad, N.I., Al-Ajlouni, M.M. and Turk, M.A. (2007) Phenotypic diversity of durum wheat (Triticum durum Desf.) from Jordan. *Genet. Resour. Crop Evol.*, 54, 129–138.
- Reilly J, Tubiello F, McCarl B, Abler D, Darwin R, Fuglie K, Hollinger S, Izaurralde C, Jagtap S, Jones J, Mearns L, Ojima D, Paul E, Paustian K, Riha S, Rosenberg N, Rosenzweig C (2003) U.S. agriculture and climate change: new results. Clim Change 57(1):43–69
- Reynolds MP, Trethowan RM. 2007. Physiological interventions in breeding for adaptation to abiotic stress. In: Scale and complexity in plant system research: Geneplant-crop relations (Eds: Spiertz JHJ, Struk PC & Van Laar HH) *Wageningen UR Fronties series Springer.* Pp. 129-146.
- Reynolds, M.P., Ortiz-Monasterio, J.I., McNab, A. 2001. Application of Physiology in Wheat Breeding.
- Reynolds, G., Payne, J., Sinun, W., Mosigil, G., & Walsh, R. P. (2011). Changes in forest land use and management in Sabah, Malaysian Borneo, 1990–2010, with a focus on the Danum Valley region. Philosophical Transactions of the Royal Society B: *Biological Sciences*, 366(1582), 3168-3176.
- Reynolds, M. P., Gutiérrez-Rodríguez, M., & Larqué-Saavedra, A. (2000). Photosynthesis of wheat in a warm, irrigated environment: I: genetic diversity and crop productivity. *Field Crops Research*, 66(1), 37-50.
- Reynolds, M., Bonnett, D., Chapman, S. C., Furbank, R. T., Manes, Y., Mather, D. E., & Parry, M. A. (2010). Raising yield potential of wheat. I. Overview of a consortium approach and breeding strategies. *Journal of experimental botany*, 62(2), 439-452.
- Reynolds, M., Foulkes, J., Furbank, R., Griffiths, S., King, J., Murchie, E., ... & Slafer, G. (2012). Achieving yield gains in wheat. *Plant, cell & environment*, 35(10), 1799-1823.
- Reynolds, M.P., Acevedo, E., Sayre, K.D. & Fischer, R.A. 1994a. Yield potential in modern wheat varieties: its association with a less competitive ideotype. *Field Crops Res.*, 37: 149-160
- Reynolds, M.P., Singh, R.P., Ibrahim, A., Ageeb, O.A.A., Larqué-Saavedra, A. & Quick, J.S. 1998a. Evaluating physiological traits to complement empirical selection for wheat in warm environments. *Euphytica*, 100: 84-95.
- Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., Sulpice, R., ... & Melchinger, A. E. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature genetics*, 44(2), 217.
- Rohlf, F.J. (1997) NTSYS-PC. Numerical taxonomy and multivariate analysis system, version 2.02e Exeter Software, New York
- Rosegrant MR, Ringler C, Sulser TB, Ewing M, Palazzo A, Zhu T, Nelson GC, Koo J, Robertson R, Msangi S, Batka M (2009). Agriculture and Food Security under Global Change: Prospects for 2025/2050. Background note for supporting the development of CGIAR Strategy and Results Framework. *International Food Policy Res.* Institute: Washington, D.C.

- Rosenzweig, C., & Tubiello, F. N. (2007). Adaptation and mitigation strategies in agriculture: an analysis of potential synergies. Mitigation and adaptation strategies for global change, 12(5), 855-873.
- Royo, C., Elias, E., Manthey, F. (2009). "Durum wheat breeding," In: *CerealsÐHandbook of plant breeding*. (Ed. Carena MJ) pp. 199±226. Springer: Berlin.
- Royo, C., Maccaferri, M., Alvaro, F., Moragues, M., Sanguineti, M.C., Tuberosa, R., Fouad, M., Luis, F.G.M., Andrea, D., Sayar, R., Miloudi, N., Nasserlehaq, N. and Dolors, V. (2010) Understanding the relationships between genetic and phenotypic structures of a collection of elite durum wheat accessions. *Field Crops Res.*, 119, 91–105.
- Ruiz, M., Giraldo, P., Royo, C., Villegas, D., Aranzana, M.J. and Carrillo, J.M. (2012) Diversity and genetic structure of a collection of Spanish durum wheat landraces. *Crop Sci.,* 52, 2262–2275.
- Sakamoto, S. and Fukui, K. (1972) Collection and preliminary observation of cultivated cereals and legumes in Ethiopia. Kyoto Univ. Afr. Stud., 7, 181–225.
- Sall, A. T., Cisse, M., Gueye, H., Kabbaj, H., Ndoye, I., Filali-Maltouf, A., ... & Bassi, F. M. (2018). Heat tolerance of durum wheat (tritcum durum desf.) elite germplasm tested along the senegal river. *Journal of Agricultural Science*; Vol. 10, No. 2.
- Scarascia Mugnozza, G.T. (2005) The contribution of italian wheat geneticists: from Nazareno Strampelli to Francesco D'Amato. Accademia Nazionale delle Scienze, Rome, Italy.
- Schiex, T., de Givry, S., Chabrier, P., Bouchez, M. (2009) Carthagène - User Documentation. INRA C. Nelson Univ. Kansas.
- Schön, C. C., Utz, H. F., Groh, S., Truberg, B., Openshaw, S., & Melchinger, A. E. (2004). Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics*, 167(1), 485-498.
- Semenov, M. A., & Halford, N. G. (2009). Identifying target traits and molecular mechanisms for wheat breeding under a changing climate. *Journal of experimental botany*, 60(10), 2791-2804.
- Semenov, M. A., Stratonovitch, P., Alghabari, F., & Gooding, M. J. (2014). Adapting wheat in Europe for climate change. *Journal of cereal science*, 59(3), 245-256.
- Sharma, S., Upadhyaya, H.D., Varshney, R.K. and Gowda, C.L.L. (2013) Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. *Front. Plant Sci.*, 4, 309.
- Shewry, P. R. (2009). Wheat. *Journal of experimental botany*, 60(6), 1537-1553.
- Singh S S, Sharma R K, Singh, Gyanendra and Saharan M S. 2011. 100 years of wheat research in India. A saga of distinguished achievements. DWR, Karnal, p 281.
- Siol, M., Jacquin, F., Chabert-Martinello, M., Smýkal, P., Le Paslier, M. C., Aubert, G., & Burstin, J. (2017). Patterns of genetic structure and linkage disequilibrium in a large collection of pea germplasm. *G3: Genes, Genomes, Genetics*, 7(8), 2461-2471.
- Smit, B., & Skinner, M. W. (2002). Adaptation options in agriculture to climate change: a typology. Mitigation and adaptation strategies for global change, 7(1), 85-114.
- Smithers, J., & Blay-Palmer, A. (2001). Technology innovation as a strategy for climate adaptation in agriculture. *Applied Geography*, 21(2), 175-197.
- Spillane, C. and Gepts, P. (2001) Evolutionary and genetic perspectives on the dynamics of crop genepools. In: Broadening the Genetic Base of Crop Production (Cooper, H.D., Spillane, C. and Hodgkin, T., eds.), pp. 25-70. Wallingford, Oxon, United Kingdom: CABI Publishing.
- Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences, 100(16), 9440-9445.
- Sukumaran, S., & Yu, J. (2014). Association mapping of genetic resources: achievements and future perspectives. In Genomics of plant genetic resources (pp. 207-235). *Springer*, Dordrecht.
- Sukumaran, S., Dreisigacker, S., Lopes, M., Chavez, P., & Reynolds, M. P. (2015). Genome-wide association study for grain yield and related traits in an elite spring wheat

population grown in temperate irrigated environments. *Theoretical and applied genetics*, 128(2), 353-363.

- Sukumaran, S., Reynolds, M. P., & Sansaloni, C. (2018). Genome-wide association analyses identify QTL hotspots for yield and component traits in durum wheat grown under yield potential, drought, and heat stress environments. *Frontiers in plant science*, 9, 81.
- Sukumaran, S., Xiang, W., Bean, S. R., Pedersen, J. F., Kresovich, S., Tuinstra, M. R., ... & Yu, J. (2012). Association mapping for grain quality in a diverse sorghum collection. *The Plant Genome*, 5(3), 126-135.
- Tahir, I. S., Nakata, N., Ali, A. M., Mustafa, H. M., Saad, A. S. I., Takata, K., ... & Abdalla, O. S. (2006). Genotypic and temperature effects on wheat grain yield and quality in a hot irrigated environment. *Plant breeding*, 125(4), 323-330.
- Tanksley, S.D. and McCouch, S.R. (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science*, 277, 1063–1066.
- Tanno, K. and Willcox, G. (2006) How fast was wild wheat domesticated? *Science*, 311, 1886.
- Tas S, Tas B. 2007. Some physiological responses of drought stress in wheat genotypes with different ploidity in Turkiye. *World J Agric Sci* 3: 178-183.
- Tester, M. and Langridge, P. (2010) Breeding technologies to increase crop production in a changing world. *Science*, 327, 818–822.
- Van Vleck, L. D., Hakim, A. F., Cundiff, L. V., Koch, R. M., Crouse, J. D., & Boldman, K. G. (1992). Estimated breeding values for meat characteristics of crossbred cattle with an animal model. *Journal of animal science*, 70(2), 363-371.
- Van Raden, P. 2008. Efficient methods to compute genomic predictions.J. Dairy Sci. 91:4414. doi:10.3168/jds.2007-0980
- Vavilov, N.I. (1951) The origin, variation, immunity and breeding of cultivated plants. trans. from the Russian by K. Starr Chester. Waltham, Mass.: Chronica Botanica. *Science*, 115, 364 pp. New York: Stechert-Hafner.
- Vavilov, N.I. (1992) Origin and geography of cultivated plants. Cambridge, United Kingdom: Cambridge University Press.
- Vijayalakshmi, K., Fritz, A. K., Paulsen, G. M., Bai, G., Pandravada, S., & Gill, B. S. (2010). Modeling and mapping QTL for senescence-related traits in winter wheat under high temperature. *Molecular Breeding*, 26(2), 163-175.
- Wang CY, Ouyang G, Ma DY, Guo TC, Zhu YJ, Cui JM, Zhang MX (2005a). Effects of sowing date on quality traits and starch pasting properties in different genotypes of wheat cultivars. *Acta Agric Boreali-Sin* 20:49-52
- Wardlaw, I. F., I. A. Dawson, P. Munibi and R. Fewster. 1989. The tolerance of wheat to high temperatures during reproductive growth. I. Survey procedures and general response patterns. *Aust. J. Agric. Res.* 40: 1- 13.
- Wientjes, Y. C., Veerkamp, R. F., & Calus, M. P. (2013). The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics*, 193(2), 621-631.
- Windhausen VS, Atlin GN, Crossa J, Hickey JM, Grudloyma P, Terekegne A et al. (2012). Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *Genes Genomes Genet* 2: 1427–1436.
- Winfield, M.O., Allen, A.M., Burridge, A,J., Barker, G,L., Benbow, H,R., Wilkinson, P.A., Coghill, J., Waterfall, C., Davassi, A., Scopes, G., Pirani, A., Webster, T., Brew, F., Bloor, C., King, J., West, C., Griffiths, S., King, I., Bentley, A,R, and Edwards K.J. (2016) High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol. J.*, 14, 1195-1206.
- Wollenweber, B., Porter, J. R., & Schellberg, J. (2003). Lack of interaction between extreme high-temperature events at vegetative and reproductive growth stages in wheat. *Journal of Agronomy and Crop Science*, 189(3), 142-150.
- Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., ... & Kresovich, S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics*, 38(2), 203.

- Zahedi M, Jenner CF. 2003. Analysis of effects in wheat of high temperature on grain filling attributes estimated from mathematical models of grain filling. *The Journal of Agricultural Science* 141: 203-212.
- Zaim, M., El Hassouni, K., Gamba, F., Filali-Maltouf, A., Belkadi, B., Sourour, A., ... & Bassi, F. M. (2017). Wide crosses of durum wheat (Triticum durum Desf.) reveal good disease resistance, yield stability, and industrial quality across Mediterranean sites. *Field crops research*, 214, 219-227.
- Zencirci, N. and Karagoz, A. (2005) Effect of developmental stages length on yield and some quality traits of Turkish durum wheat (Triticum turgidum L. convar. durum (Desf.) Mackey) landraces: Influence of developmental stages length on yield and quality of durum wheat. *Genet. Resour. Crop Evol.*, 52, 765–774.
- Zhang, X., Chen, S., Sun, H., Wang, Y., & Shao, L. (2010). Water use efficiency and associated traits in winter wheat cultivars in the North China Plain. *Agricultural Water Management,* 97(8), 1117-1125.
- Zhang, Y. Q., Deng, Y., Chen, R. Y., Cui, Z. L., Chen, X. P., Yost, R., ... & Zou, C. Q. (2012). The reduction in zinc concentration of wheat grain upon increased phosphorus-fertilization and its mitigation by foliar zinc application. *Plant and Soil*, 361(1-2), 143-152.
- Zhao, Y., Zeng, J., Fernando, R., & Reif, J. C. (2013). Genomic prediction of hybrid wheat performance. *Crop Science*, 53(3), 802-810.
- Zheng, R. L., Cai, C., Liang, J. H., Huang, Q., Chen, Z., Huang, Y. Z., ... & Sun, G. X. (2012). The effects of biochars from rice residue on the formation of iron plaque and the accumulation of Cd, Zn, Pb, As in rice (Oryza sativa L.) seedlings. *Chemosphere*, 89(7), 856-862.
- Zhong, S., J.C.M. Dekkers, R.L. Fernando, and J.L. Jannink. 2009. Factors aff ecting accuracy from genomic selection in populations derived from multiple inbred lines: A barley case study. *Genetics* 182:355. doi:10.1534/genetics.108.098277
- Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and prospects of association mapping in plants. *The plant genome*, 1(1), 5-20.
- Zohary, D., Hopf, M. and Weiss, E. (2012) The domestication of the Plants in the Old World: the origin and spread of cultivated plants in West Asia, Europe and Nile valley. Oxford, United Kingdom: Oxford University Press.

Annexes

Annex 1: Complete list with pedigree, origin, IG, and sub-population assignment for each software

DurA M Core	Source	ID (ig)	Origin	Pedigree	Genotyped	Assignment	grp DAPC	clusters DAPC	grp structure	Clusters structure	grp NJ TREE	clusters NJ
5	DWGIGSSALATFIN A036	92324	SYRIA	SYR833::55	yes	solid	1	middle east	1,2,4	Landraces	1	Landraces 1
7	DWGIGSSALATFIN A048	96252	JORDAN	JOR832::46	yes	solid	1	middle east	1,2,4	Landraces	1	Landraces 1
8	DWGIGSSALATFIN A053	97186	JORDAN	JOR832::49	yes	solid	1	middle east	1,2,4	Landraces	1	Landraces 1
11	DW-SR-FIGS046	82452	IRAQ	IRQS167	yes	solid	1	middle east	1,2,4	Landraces	1	Landraces 1
45	FIGSDRYWET098	96259	JORDAN	JOR84::131	yes	solid	1	middle east	1,2,4	Landraces	1	Landraces 1
46	FIGSDRYWET099	97225	JORDAN	JOR84::132	yes	solid	1	middle east	1,2,4	Landraces	1	Landraces 1
54	FIGSDRYWET130	99261	SYRIA	SYR923:8	yes	solid	1	middle east	1,2,4	Landraces	1	Landraces 1
86	Syr-36	95884	SYRIA	TallHasil	yes	solid	1	middle east	1,2,4	Landraces	1	Landraces 1
87	Syr-60	95925	SYRIA	Jnan	yes	solid	1	middle east	1,2,4	Landraces	1	Landraces 1
88	Syr-67	96149	SYRIA	Mansura;12kmSouth	yes	solid	1	middle east	1,2,4	Landraces	1	Landraces 1
89	Syr-72	96156	SYRIA	30kmSEKa'aLuly	yes	solid	1	middle east	1,2,4	Landraces	1	Landraces 1
10	DW-SR-FIGS027	76851	ETHIOPIA	ETH74::13	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
22	FIGSDRYWET004	79533	ETHIOPIA	ETH74::53	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
25	FIGSDRYWET014	82191	RUSSIA	RUSS200	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
29	FIGSDRYWET028	83807	ETHIOPIA	ETH64:112	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
30	FIGSDRYWET029	83842	ETHIOPIA	ETH64:151	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
31	FIGSDRYWET030	83846	ETHIOPIA	ETH64:157	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
39	FIGSDRYWET058	85981	ETHIOPIA	ETH64:17	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
42	FIGSDRYWET086	90460	ETHIOPIA	ETH731::84	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
43	FIGSDRYWET091	90482	ETHIOPIA	ETH731::112	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
44	FIGSDRYWET096	96186	JORDAN	JOR882:48	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
51	FIGSDRYWET110	99205	YEMEN	YEM801:2	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3

64	FIGSDWHOTCLD009	79653	ETHIOPIA	ETH74::56	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
72	FIGSDWHOTCLD093	88921	ETHIOPIA	ETH64:73	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
73	FIGSDWHOTCLD107	90410	ETHIOPIA	ETH731::43	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
74	FIGSDWHOTCLD122	90429	ETHIOPIA	ETH731::48	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
75	FIGSDWHOTCLD133	90458	ETHIOPIA	ETH731::89	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
76	FIGSDWHOTCLD134	90459	ETHIOPIA	ETH731::88	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
77	FIGSDWHOTCLD139	90660	ETHIOPIA	ETH74::23	yes	solid	2	T.abysinicum type	1,2,4	Landraces	3	landraces 3
37	FIGSDRYWET051	85521	AFGHANSTAN	AFG54::17	yes	solid	3	Mediterranean trades	3	Old italian variety	2	Landraces 2
41	FIGSDRYWET078	90233	AFGHANSTAN	AFG68::77	yes	solid	3	Mediterranean trades	3	Old italian variety	2	Landraces 2
49	FIGSDRYWET108	98797	IRAN	IRNS294	yes	solid	3	Mediterranean trades	3	Old italian variety	2	Landraces 2
94	Tun-52	95200	TUNISIA	E102242	yes	solid	3	Mediterranean trades	3	Old italian variety	2	Landraces 2
95	Alg-03		ALGERIA	DZA	yes	solid	3	Mediterranean trades	3	Old italian variety	2	Landraces 2
96	Alg-15		ALGERIA	DZA	yes	solid	3	Mediterranean trades	3	Old italian variety	2	Landraces 2
34	FIGSDRYWET038	85026	SPAIN	ESP::29	yes	solid	3	Mediterranean trades	3	Old italian variety	2	Landraces 2
80	FIGSDWHOTCLD149	90743	ITALY	ITA73::26	yes	solid	3	Mediterranean trades	3	Old italian variety	2	Landraces 2
83	FIGSDWHOTCLD168	97574	ETHIOPIA	ETH50/51::10	yes	solid	3	Mediterranean trades	3	Old italian variety	2	Landraces 2
90	Tun-14	94128	TUNISIA	E0082450	yes	solid	3	Mediterranean trades	3	Old italian variety	2	Landraces 2
91	Tun-19	94302	TUNISIA	E0836	yes	solid	3	Mediterranean trades	3	Old italian variety	2	Landraces 2
18	DW-SR-FIGS103	93114	ALGERIA	DZA75::101	yes	good	3	Mediterranean trades	3	Old italian variety	4	Cultivars 1
66	FIGSDWHOTCLD021	82109	AZERBIJAN	AZES632	yes	good	3	Mediterranean trades	3	Old italian variety	4	Cultivars 1
82	FIGSDWHOTCLD166	97150	GREECE	GRC78::45	yes	good	3	Mediterranean trades	3	Old italian variety	4	Cultivars 1
92	Tun-29	94615	TUNISIA	E0832	yes	good	3	Mediterranean trades	3	Old italian variety	4	Cultivars 1
93	Tun-37	94806	TUNISIA	E0105257	yes	good	3	Mediterranean trades	3	Old italian variety	4	Cultivars 1
1	DWGIGSSALATFINA019	85846	SPAIN	ESPS1603	yes	good	3	Mediterranean trades	3	Old italian variety	5	Cultivars 2

47	FIGSDRYWET105	98680	CHAINA	CHN::25	yes	good	3	Mediterranean trades	1,2,4	Landraces	2	Landraces 2
71	FIGSDWHOTCLD090	88293	YUGOSLAVIA	YUG56::14	yes	good	3	Mediterranean trades	1,2,4	Landraces	2	Landraces 2
343	DURUM_PANEL_UNIBO-0250	Ardente	ITALY	ISRAEL-DUM-303/PRELIMINARY-77//664	yes	bad	3	Mediterranean trades	7	breeding programs exchange	4	Cultivars 1
28	FIGSDRYWET023	83202	KAZAKHSTAN	KAZS31	yes	bad	3	Mediterranean trades	8	Dev countries	4	Cultivars 1
6	DWGIGSSALATFINA038	93139	ALGERIA	DZA75::95	yes	bad	3	Mediterranean trades	1,2,4	Landraces	4	Cultivars 1
9	DWGIGSSALATFINA057	97548	SPAIN	ESPS1930	yes	bad	3	Mediterranean trades	1,2,4	Landraces	4	Cultivars 1
14	DW-SR-FIGS082	88631	GREECE	BALK42::151	yes	bad	3	Mediterranean trades	1,2,4	Landraces	4	Cultivars 1
23	FIGSDRYWET007	81953	RUSSIA	RUSS170	yes	bad	3	Mediterranean trades	1,2,4	Landraces	4	Cultivars 1
68	FIGSDWHOTCLD063	85015	SPAIN	ESPS1919	yes	bad	3	Mediterranean trades	1,2,4	Landraces	4	Cultivars 1
53	FIGSDRYWET120	99224	YEMEN	YEM801:69	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
27	FIGSDRYWET022	83103	GEORGIA	GEOS58	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
52	FIGSDRYWET115	99214	YEMEN	YEM801:12	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
2	DWGIGSSALATFINA020	86075	INDIA	INDS413	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
3	DWGIGSSALATFINA031	89459	AFGHANSTAN	AFG68::50	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
12	DW-SR-FIGS049	82578	AFGHANSTAN	AFG37::5	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
15	DW-SR-FIGS089	90229	AFGHANSTAN	AFG68::74	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
26	FIGSDRYWET016	82244	RUSSIA	RUSS199	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
32	FIGSDRYWET033	84294	TURKY	TUR48D:6	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
33	FIGSDRYWET036	84832	INDIA	IND47/48::47	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
55	FIGSDRYWET138	113071	SAU	SAU89:7	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
56	FIGSDRYWET143	118752	IRAN	IRNS375	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
57	FIGSDRYWET144	119026	IRAN	IRNS382	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
58	FIGSDRYWET147	126364	ARMINIA	ARM99:28	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1

59	FIGSDRYWET160	127271	KAZAKHSTAN	KAZ28::59	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
60	FIGSDRYWET163	127276	RUSSIA	RUSS2843	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
61	FIGSDRYWET174	130921	ARMINIA	ARM25::33	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
62	FIGSDRYWET179	131004	PAKISTAN	PAK28::27	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
63	FIGSDWHOTCLD005	43305	OMAN	OMN87:70	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
79	FIGSDWHOTCLD147	90738	ITALY	ITA73::22	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
84	FIGSDWHOTCLD182	130807	ITALY	ITA27::67	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
85	FIGSDWHOTCLD188	130987	PAKISTAN	PAKS237	yes	solid	4	Central and South Asian landraces	1,2,4	Landraces	1	Landraces 1
13	DW-SR-FIGS063	84819	IRAQ	IRQS149	yes	good	4	Central and South Asian landraces	1,2,4	Landraces	3	landraces 3
38	FIGSDRYWET053	85620	AFGHANSTAN	KUSE::1	yes	good	4	Central and South Asian landraces	1,2,4	Landraces	3	landraces 3
50	FIGSDRYWET109	99022	TURKY	TUR::1	yes	good	4	Central and South Asian landraces	1,2,4	Landraces	3	landraces 3
48	FIGSDRYWET107	98749	KAZAKHSTAN	KAZ::18	yes	good	4	Central and South Asian landraces	8	Dev countries	1	Landraces 1
36	FIGSDRYWET048	85502	INDIA	INDS247	yes	bad	4	Central and South Asian landraces	8	Dev countries	3	landraces 3
4	DWGIGSSALATFINA035	90731	ITALY	ITA73::15	yes	solid	5	Om Rabi derivatives	5	Omrabi deriv	4	Cultivars 1
176	DURUM_PANEL_UNIBO-060		ICARDA	AW12/BIT	yes	solid	5	Om Rabi derivatives	5	Omrabi deriv	4	Cultivars 1
210	DURUM_PANEL_UNIBO-0160	MASSARA1	ICARDA	Mrb3/4/BYE*2/Tc//ZB/W/3/Cit	yes	solid	5	Om Rabi derivatives	5	Omrabi deriv	4	Cultivars 1
237	MCHCB-0176	Zeina8	ICARDA	GdoVZ512/Cit//Ruff/Fg/3/Src3	yes	solid	5	Om Rabi derivatives	5	Omrabi deriv	4	Cultivars 1
322	DURUM_PANEL_UNIBO-077	APPIO	ITALY	CAPELLI//GAVIOTA/YUMA	yes	solid	5	Om Rabi derivatives	5	Omrabi deriv	4	Cultivars 1
324	DURUM_PANEL_UNIBO-083	CAPEITI8	ITALY	EITI*6/SENATORE-CAPPELLI	yes	solid	5	Om Rabi derivatives	5	Omrabi deriv	4	Cultivars 1
351	DURUM_PANEL_UNIBO-058	TOMOUH	MOROCCO-TUNISIA	Jori c69/Hau	yes	solid	5	Om Rabi derivatives	5	Omrabi deriv	4	Cultivars 1
381	DURUM_MAS	Omrabi 5	Syria	Jori c69/Hau	yes	solid	5	Om Rabi derivatives	5	Omrabi deriv	4	Cultivars 1
383	DURUM_MAS	Icarasha2		Stj3//Bcr/Lks4/3/Ter3	yes	solid	5	Om Rabi derivatives	5	Omrabi deriv	4	Cultivars 1
211	DURUM_PANEL_UNIBO-0162	Omrabi 17	ICARDA	Jori c69/Hau	yes	solid	5	Om Rabi derivatives	5	Omrabi deriv	4	Cultivars 1

213	DURUM_PANEL_UNIBO-0165	OMBIT1	ICARDA	Omrabi5/Albit1	yes	solid	5	Om Rabi derivatives	5	Omrabi deriv	4	Cultivars 1
229	MCHCB-0124		ICARDA	Stk/Hau//Heca1	yes	solid	5	Om Rabi derivatives	5	Omrabi deriv	4	Cultivars 1
323	DURUM_PANEL_UNIBO-079	Arcangelo	ITALY	CRESO/APPULO	yes	good	5	Om Rabi derivatives	3	Old italian variety	4	Cultivars 1
113	42-DCC-0321	Po	ICARDA	not available	yes	solid	6	Italian breeding program	6	italian varieties	5	Cultivars 2
248	Jable-007	Hyperno	AUSTRALIA	KALKA(SIB)/TAMAROI	yes	solid	6	Italian breeding program	6	italian varieties	5	Cultivars 2
307	DURUM_PANEL_UNIBO-0192	Aramon	FRANCE	not available	yes	solid	6	Italian breeding program	6	italian varieties	5	Cultivars 2
314	DURUM_PANEL_UNIBO-01	BRADANO	ITALY	not available	yes	solid	6	Italian breeding program	6	italian varieties	5	Cultivars 2
315	DURUM_PANEL_UNIBO-02	CANNIZZO	ITALY	CAPEITI/VALNOVA(F5)//(F5)PATRIZIO/VALFORTE	yes	solid	6	Italian breeding program	6	italian varieties	5	Cultivars 2
316	DURUM_PANEL_UNIBO-03	CLAUDIO	ITALY	SEL.CIMMYT-35/DURANGO//ISEA-1938/GRAZIA	yes	solid	6	Italian breeding program	6	italian varieties	5	Cultivars 2
318	DURUM_PANEL_UNIBO-06	Mongibello	ITALY	TRINAKRIA/VALFORTE	yes	solid	6	Italian breeding program	6	italian varieties	5	Cultivars 2
330	DURUM_PANEL_UNIBO-094	FORTORE	ITALY	CAPEITI-8/VALFORTE	yes	solid	6	Italian breeding program	6	italian varieties	5	Cultivars 2
331	DURUM_PANEL_UNIBO-096	GRAZIA	ITALY	M-6800127(ISWRN-21)/VALSELVA	yes	solid	6	Italian breeding program	6	italian varieties	5	Cultivars 2
335	DURUM_PANEL_UNIBO-0105	OFANTO	ITALY	ADAMELLO/APPULO	yes	solid	6	Italian breeding program	6	italian varieties	5	Cultivars 2
340	DURUM_PANEL_UNIBO-0114	VALNOVA	ITALY	GIORGIO-324//SENATORE-CAPELLI/YUMA	yes	solid	6	Italian breeding program	6	italian varieties	5	Cultivars 2
341	DURUM_PANEL_UNIBO-0115	VARANO	ITALY	CAPEITI-8/CRESO//CRESO/3/VALFORTE/TRINAKRIA	yes	solid	6	Italian breeding program	6	italian varieties	5	Cultivars 2
376	DURUM_PANEL_UNIBO-076	ANTON	SPAIN-CIMMYT	not available	yes	solid	6	Italian breeding program	6	italian varieties	5	Cultivars 2
320	DURUM_PANEL_UNIBO-09	QUADRATO	ITALY	CRESO/TRINAKRIA	yes	solid	6	Italian breeding program	6	italian varieties	5	Cultivars 2
321	DURUM_PANEL_UNIBO-010	Torrebianca	ITALY	CRESO/TRINAKRIA	yes	solid	6	Italian breeding program	6	italian varieties	5	Cultivars 2
219	DURUM_PANEL_UNIBO-0182	TELSET5	ICARDA	SD 8036/Mtl-1//Albit3	yes	good	6	Italian breeding program	5	Omrabi deriv	5	Cultivars 2
336	DURUM_PANEL_UNIBO-0106	PLATANI	ITALY	VALNOVA/CAPEITI-8	yes	good	6	Italian breeding program	5	Omrabi deriv	5	Cultivars 2
312	DURUM_PANEL_UNIBO-0205	Nefer	FRANCE	164/KEOPS	yes	good	6	Italian breeding program	7	breeding programs exchange	5	Cultivars 2

253	DURUM_PANEL_UNIBO-0244	Kamilaroi	AUSTRALIA	DURATI(SIB)/LEEDS	yes	good	6	Italian breeding program	3	Old italian variety	5	Cultivars 2
319	DURUM_PANEL_UNIBO-08	Pietrafritta	ITALY	GRAZIA/ISA	yes	good	6	Italian breeding program	6	italian varieties	4	Cultivars 1
35	FIGSDRYWET043	85403	ETHIOPIA	ETHS1134	yes	bad	6	Italian breeding program	3	Old italian variety	4	Cultivars 1
40	FIGSDRYWET060	85991	ETHIOPIA	ETH64:47	yes	bad	6	Italian breeding program	3	Old italian variety	4	Cultivars 1
69	FIGSDWHOTCLD065	85222	ETHIOPIA	ETHS948	yes	bad	6	Italian breeding program	3	Old italian variety	4	Cultivars 1
70	FIGSDWHOTCLD073	85404	ETHIOPIA	ETH50/51::21	yes	bad	6	Italian breeding program	3	Old italian variety	4	Cultivars 1
78	FIGSDWHOTCLD144	90724	ITALY	ITA73::9	yes	bad	6	Italian breeding program	3	Old italian variety	4	Cultivars 1
81	FIGSDWHOTCLD150	90744	ITALY	ITA73::27	yes	bad	6	Italian breeding program	3	Old italian variety	4	Cultivars 1
298	DURUM_PANEL_UNIBO-0264		CIMMYT	RANCO//CIT71/CII/3/COMDK/4/TCHO//SHWA/MALD/3/CREX/5/SNTURKMI8384375/NIGRIS_5//TANTLO_1	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
301	DURUM_PANEL_UNIBO-0268		CIMMYT	ARMENT//SRN_3/NIGRIS_4/3/CANELO_9.1	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
300	DURUM_PANEL_UNIBO-0267		CIMMYT	ROLA_5/3/AJAIA_12/F3LOCAL(SEL.ETHIO.135.85)//PLATA_13/4/MALMUK_1/SERRATOR_1	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
273	DURUM_PANEL_UNIBO-013		CIMMYT	CIMMYT41_(DUKEM/3/RUFF/FGO//YAV79)	yes	good	7	breeding programs exchange	7	breeding programs exchange	4	Cultivars 1
286	DURUM_PANEL_UNIBO-0103	MEXICALI75	CIMMYT	GERARDO-VZ-469/3/JORI(SIB)//ND-61-130/LEEDS	yes	good	7	breeding programs exchange	6	ICARDA wide	5	Cultivars 2
289	DURUM_PANEL_UNIBO-0254	Samaypa C2004	CIMMYT	SOMAT_4/INTER_8(=SAMAYOAC2004)	yes	good	7	breeding programs exchange	9	ICARDA	5	Cultivars 2
304	FIGSDWHOTCLD089	88029	ETHIOPIA	ETHS1035	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
305	FIGSDWHOTCLD170	97901	FRANCE	FRAS129	yes	good	7	breeding programs exchange	7	breeding programs exchange	4	Cultivars 1
308	DURUM_PANEL_UNIBO-0193	Arcalis	FRANCE	EDMORE/CRESO	yes	good	7	breeding programs exchange	3	Old italian variety	5	Cultivars 2

306	DURUM_PANEL_UNIBO-0191	Agridur	FRANCE	EDMORE//CIMMYT-303/CHANDUR	yes	good	7	breeding programs exchange	8	Dev countries	5	Cultivars 2
310	DURUM_PANEL_UNIBO-0200	Orjaune	FRANCE	MIRADUR/IDYN-81-04	yes	good	7	breeding programs exchange	8	Dev countries	5	Cultivars 2
236	MCHCB-0169		ICARDA	((SwAlg/Gd1)81/Ch1)46	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
196	DURUM_PANEL_UNIBO-0137	Deraa	ICARDA	Can2101/Magh//Stk/3/Wlls/65150	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
181	DURUM_PANEL_UNIBO-066		ICARDA	KRS/HAUCAN_(KRS/HCN)	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
185	DURUM_PANEL_UNIBO-073	SEBAH	ICARDA	Sbu/Lahn	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
202	DURUM_PANEL_UNIBO-0147	SEBOU	ICARDA	Cr/T.Polonicum	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
239	MCHCB-0210		ICARDA	Azeghar1/HFN94N75//Vitron/Bicrederaa1	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
215	DURUM_PANEL_UNIBO-0172	OUASERL1	ICARDA	Ossl1/4/MrbSH/3/Rabi//Gs/Cr/5/Krs/Hcn	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
200	DURUM_PANEL_UNIBO-0145	HEIDER	ICARDA	Can2109//Jo/AA/3/S15/Cr	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
173	IDYT37-21	Bezaghras	ICARDA	Ossl1/Stj5/5/Bicrederaa1/4/BEZAIZSHF//SD19539/Waha/3/Stj/Mrb3/6/Mgnl3/Aghrass2	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
186	DURUM_PANEL_UNIBO-074	STOJOCRI3	ICARDA	Stk/4/Jo/3/Jo/Cr//Cit71	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
217	DURUM_PANEL_UNIBO-0180	SEBATEL1	ICARDA	Sbh/4/D68193A1A//Ruff/Fg/3/Mtl5	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
223	MCHCB-082		ICARDA	Bicrederaa-1/Tavoliere//Gdr1	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
192	DURUM_PANEL_UNIBO-0126	AWALI1	ICARDA	Cit//D.dwarfS15/Cr	yes	good	7	breeding programs exchange	7	breeding programs exchange	6	Cultivars 3
178	DURUM_PANEL_UNIBO-062		ICARDA	CHHB88/DERAA	yes	good	7	breeding programs exchange	7	breeding programs exchange	6	Cultivars 3

112	DWAyT-0322		ICARDA	Heider/TAraticumMA//Mrb5	yes	good	7	breeding programs exchange	7	breeding programs exchange	6	Cultivars 3
226	MCHCB-0102		ICARDA	OmRabi3/T.urartu500651/Ch5//980947/3/Otb4//Ossl1/Rfm6	yes	good	7	breeding programs exchange	7	breeding programs exchange	6	Cultivars 3
197	DURUM_PANEL_UNIBO-0138	FURAT 1	ICARDA	Snipe/3/Jo/Cr//Gs/AA	yes	good	7	breeding programs exchange	7	breeding programs exchange	4	Cultivars 1
198	DURUM_PANEL_UNIBO-0141	GIDAR A2	ICARDA	Stj/Mrb3	yes	good	7	breeding programs exchange	5	Omrabi deriv	5	Cultivars 2
206	DURUM_PANEL_UNIBO-0154	Korifla = Cham 3	ICARDA	DS15/Geier	yes	good	7	breeding programs exchange	6	ICARDA wide	5	Cultivars 2
188	DURUM_PANEL_UNIBO-0120	ANGRE	ICARDA	not available	yes	good	7	breeding programs exchange	6	ICARDA wide	5	Cultivars 2
234	MCHCB-0158		ICARDA	(Jk/Ch1)-506	yes	good	7	breeding programs exchange	8	Dev countries	5	Cultivars 2
238	MCHCB-0180		ICARDA	OmRabi3/T.urartu500651//ICAMORTA0463/Ammar8/4/Stj3//Bcr/Lks4/3/Ter3	yes	good	7	breeding programs exchange	1,2,4	Landraces	5	Cultivars 2
203	DURUM_PANEL_UNIBO-0150	JORDAN	ICARDA	GdoVZ469/Plc/5/21563/3/LK/Ld390//Ch67/4/Cit71	yes	bad	7	breeding programs exchange	3	Old italian variety	4	Cultivars 1
334	DURUM_PANEL_UNIBO-0102	MESSAPIA	ITALY	MEXA/(SIB)CRANE//TITO	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
339	DURUM_PANEL_UNIBO-0111	SVEVO	ITALY	CIMMYT-SELECTION/ZENIT	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
338	DURUM_PANEL_UNIBO-0108	PRODURA	ITALY	TREMES-MOLLE-ENANO(DWARF)/2*TEHUACAN-60/3/ZENATI-BOUTEILLE/WELLS/4/2*BARRIGON-YAQUI-ENANO(DWARF)/TEHUACAN-60//TACUR-TIPO-125-E/2*TEHUACAN-60	yes	good	7	breeding programs exchange	7	breeding programs exchange	4	Cultivars 1
325	DURUM_PANEL_UNIBO-087	COLOSSEO	ITALY	CRESO/(M)MEXA[	yes	good	7	breeding programs exchange	7	breeding programs exchange	4	Cultivars 1
326	DURUM_PANEL_UNIBO-089	CRESO	ITALY	CAPELLI-B-144/5/YAKTANA-54//(SELECTION-	yes	good	7	breeding programs exchange	7	breeding programs exchange	4	Cultivars 1

				14)NORIN-10/BREVOR/3/CAPELLI-63/4/3*TEHUACAN-60								
333	DURUM_PANEL_UNIBO-0101	LIRA_B_45	ITALY	MANDON/FD-1104	yes	good	7	breeding programs exchange	3	Old italian variety	5	Cultivars 2
342	DURUM_PANEL_UNIBO-0249	Levante	ITALY	G-80/PICENO//IONIO	yes	good	7	breeding programs exchange	8	Dev countries	5	Cultivars 2
348	DURUM_PANEL_UNIBO-050	MAROUANE	MOROCCO	INRAM.1809	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
346	DURUM_PANEL_UNIBO-048	CHAOUI	MOROCCO	INRA1807;INRA1807	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
347	DURUM_PANEL_UNIBO-049	AMRIA	MOROCCO	INRAM.1808	yes	good	7	breeding programs exchange	7	breeding programs exchange	4	Cultivars 1
350	DURUM_PANEL_UNIBO-057	TAREK	MOROCCO-TUNISIA	YAVAROS/SAPI//YAVAROS-79/3/STU	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
353	DURUM_PANEL_UNIBO-052	ISLY	MOROCCO-TUNISIA, INRA MAROC_Mar_1988	ERPEL(SIB)/(SIB)RUSO	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
357	DURUM_PANEL_UNIBO-055	MARZAK	MOROCCO-TUNISIA,INRA MAROC_Mar_1984	not available	yes	good	7	breeding programs exchange	7	breeding programs exchange	4	Cultivars 1
361	DURUM_PANEL_UNIBO-0100	KRONOS	SOUTH-WESTERN USA	Male_Sterile_Facilitated_Recurrent_Selection_Population-APB-MSFRS-Pop	yes	good	7	breeding programs exchange	6	ICARDA wide	5	Cultivars 2
362	DURUM_PANEL_UNIBO-0104	MOHAWK	SOUTH-WESTERN USA	(S)Male_Sterile_Facilitated_Recurrent_Selection_Population-883-22-ALPHA-85-CHA	yes	good	7	breeding programs exchange	8	Dev countries	5	Cultivars 2
363	DURUM_PANEL_UNIBO-0109	REVA	SOUTH-WESTERN USA	(S)Male_Sterile_Facilitated_Recurrent_Selection_Population-AZ-85-MSFRS-Durum-Population	yes	good	7	breeding programs exchange	9	ICARDA	5	Cultivars 2

369	DURUM_PANEL_UNIBO-034	BOLIDO	SPAIN	not available	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
367	DURUM_PANEL_UNIBO-032	BOABDIL	SPAIN	not available	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
372	DURUM_PANEL_UNIBO-039	DURCAL	SPAIN	not available	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
368	DURUM_PANEL_UNIBO-033	BOLENGA	SPAIN	not available	yes	good	7	breeding programs exchange	8	Dev countries	5	Cultivars 2
373	DURUM_PANEL_UNIBO-044	SENADUR	SPAIN	not available	yes	bad	7	breeding programs exchange	8	Dev countries	6	Cultivars 3
377	DURUM_PANEL_UNIBO-0110	ROQUENO	SPAIN-CIMMYT	not available	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
380	DURUM_PANEL_UNIBO-0241	Tacna	USA-GREAT-PLAINS	Male_Sterile_Facilitated_Recurrent_Selection_Population-Durum-S-1-E.Wheat-89-S-1	yes	good	7	breeding programs exchange	6	ICARDA wide	5	Cultivars 2
384	DURUM_MAS	Azeghar2	ICARDA	20048 Traikia(Mor)/Mrb5//Stj3	yes	solid	7	breeding programs exchange	7	breeding programs exchange	5	Cultivars 2
16	DW-SR-FIGS096	92871	ALGERIA	DZA75::35	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
17	DW-SR-FIGS097	92879	ALGERIA	DZA75::35	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
252	DURUM_PANEL_UNIBO-0243	Yallaroi	AUSTRALIA	GUILLEMOT(SEL.3)/(SIB)KAMILAROI	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
255	DURUM_PANEL_UNIBO-0208	Extradur	AUSTRALIA	GRANDUR/MONDUR//ASTRODUR	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
256	DURUM_PANEL_UNIBO-0209	Goldur	AUSTRALIA	PANDUR/VALGERARDO//VALDUR	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
257	DURUM_PANEL_UNIBO-0211	Frankodur	AUSTRALIA	MONDUR/GRANDUR//ASTRODUR	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
258	DURUM_PANEL_UNIBO-0213	Semperdur	AUSTRALIA	ASTRODUR/KAMILAROI	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
259	DURUM_PANEL_UNIBO-0217	AC Morse	CANADA	RL 7196/D84328	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
260	DURUM_PANEL_UNIBO-0218	AC Navigator	CANADA	Kyle/WB881	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
261	DURUM_PANEL_UNIBO-0219	AC Pathfinder	CANADA	Dt367/WB881	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3

262	DURUM_PANEL_UNIBO-0220	Hercules	CANADA	RL-3097/RL-3304//STEWART(TR.DR)/RL-3380	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
263	DURUM_PANEL_UNIBO-0221	Kyle	CANADA	Wakooma/Dt320//Wakooma/Dt322	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
264	DURUM_PANEL_UNIBO-0222	Medora	CANADA	Ward/Macoun	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
265	Cpozniak-003	CDC Desire	CANADA	not available	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
266	Cpozniak-004	CDC Vivid	CANADA	not available	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
267	Cpozniak-005	DT570	CANADA	not available	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
268	Cpozniak-008	Commander	CANADA	W-9260-BK-03/AC-NAVIGATOR//AC-PATHIFINDER	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
269	Cpozniak-015	AAC Raymore	CANADA	9675-AP-2/DT-732//STRONGFIELD	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
270	Cpozniak-016	Napoleon	CANADA	VIC/DT-384//DT-471	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
272	DURUM_PANEL_UNIBO-012		CIMMYT	CIMMYT36_(CMH82A.1062/3/GG0VZ394//SBA81/PLC/4/AAZ…)	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
309	DURUM_PANEL_UNIBO-0199	Galadur	FRANCE	EDMORE/BLONDUR//MONTFERRIER/DT-192	yes	good	8	Developed countries	8	Dev countries	5	Cultivars 2
311	DURUM_PANEL_UNIBO-0202	Tetradur	FRANCE	EDMORE//CAPDUR/REGAL	yes	good	8	Developed countries	8	Dev countries	5	Cultivars 2
218	DURUM_PANEL_UNIBO-0181	SHABHA	ICARDA	Landrace selection	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
332	DURUM_PANEL_UNIBO-098	ITALO	ITALY	TURCHIA//CRESO/CAPEITI-8	yes	good	8	Developed countries	8	Dev countries	5	Cultivars 2
359	DURUM_PANEL_UNIBO-081	BRAVADUR	SOUTH-WESTERN USA	(S)Male_sterile_facilitated_recurrent_selection_population-MSFRS-HRS-Quality-Population	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
360	DURUM_PANEL_UNIBO-092	DUREX	SOUTH-WESTERN USA	(S)MALE_STERILE_FACILITATED_RECURRENT_SELECTION_POPULATION_AZ-MFSRS-86	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
364	DURUM_PANEL_UNIBO-0116	WEST BRED 881	SOUTH-WESTERN USA	not available	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3

365	DURUM_PANEL_UNIBO-0189	KOFA	SOUTH-WESTERN USA	(S)Male_Sterile_Facilitated_Recurrent_Selection_Population-DICOCCUM-ALPHA-POP-85-S-1	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
374	DURUM_PANEL_UNIBO-030	ARTENA	SPAIN, SEMENCES DE PROVENCES_Fra_1993	not available	yes	good	8	Developed countries	8	Dev countries	5	Cultivars 2
379	DURUM_PANEL_UNIBO-0232	Munich	USA-GREAT-PLAINS	D-8030/D-8016	yes	solid	8	Developed countries	8	Dev countries	6	Cultivars 3
254	DURUM_PANEL_UNIBO-0245	Wollaroi	AUSTRALIA	TAM-1-B-17/(SIB)KAMILAROI//ROKEL(SIB)/(SIB)KAMILAROI**[113][1335]**;DURATI(SIB)/LEEDS	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
295	DURUM_PANEL_UNIBO-0261		CIMMYT	CNDO/PRIMADUR//HAIOU_17/3/SNTURKMI8384375/NIGRIS_5//TANTLO_1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
303	DURUM_PANEL_UNIBO-0270	Yavaros 79 = Karim = Bittern	CIMMYT	JORI69(SIB)/(SIB)ANHINGA//(SIB)FLAMINGO,MEX	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
225	MCHCB-095		ICARDA	Mck2/Tilo2//Bcrch1/Kund1149	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
240	MCHCB-0213		ICARDA	Icasyr1/4/Assassa//Waha/Brch/3/Bicredera a1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
214	DURUM_PANEL_UNIBO-0171	Outrob6	ICARDA	Ouassel1/4/GdoVZ512/Cit//Ruff/Fg/3/Pin/Gre//Trob	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
97	DAWRyT-0106		ICARDA	Amedakul1/TdicoSyrCol//Loukos	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
98	DAWRyT-0110		ICARDA	Amedakul1/TdicoSyrCol//Cham1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
99	DAWRyT-0208		ICARDA	Korifla/AegSpeltoides Syr/Amedakul	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
101	DAWRyT-0308		ICARDA	Korifla/AegSpeltoides Syr//Heider	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
102	DAWRyT-0315		ICARDA	Korifla/AegSpeltoides Syr//Mrb5	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
104	DWAyT-0205		ICARDA	Younes/TdicoAlpCol//Korifla	yes	solid	9	ICARDA breeding program	9	ICARDA wide	5	Cultivars 2

105	DWAyT-0209		ICARDA	Korifla/AegSpeltoides Syr//Amedakul	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
106	DWAyT-0212		ICARDA	Korifla/AegSpeltoides Syr//Amedakul	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
107	DWAyT-0214		ICARDA	Korifla/AegSpeltoides Syr//Amedakul	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
108	DWAyT-0215		ICARDA	Korifla/AegSpeltoides Syr//Amedakul	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
109	DWAyT-0217		ICARDA	Korifla/AegSpeltoides Syr//Loukos	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
110	DWAyT-0224		ICARDA	Korifla/AegSpeltoides Syr//Waha	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
111	DWAyT-0306		ICARDA	Korifla/AegSpeltoides Syr//Heider	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
114	ADYT_008		ICARDA	Ossl1/Stj5/5/Bidra1/4/BezaizSHF//SD19539/Waha/3/Stj/Mrb3/6/Icajihan1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
115	ADYT_009		ICARDA	Azeghar1/6/Zna1/5/Awl1/4/Ruff//Jo/Cr/3/F9.3/7/Azeghar1//Msbl1/Quarmal	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
116	ADYT_018		ICARDA	Ter1//Mrf1/Stj2/3/Icasyr1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
117	ADYT_019		ICARDA	Ter1//Mrf1/Stj2/3/Icasyr1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
120	ADYT_097		ICARDA	Ossl1/Stj5/5/Bicrederaa1/4/BEZAIZSHF//SD19539/Waha/3/Stj/Mrb3/6/Stk/Hau//Heca1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
121	ADYT_104		ICARDA	Bcr/Lks4//Mrf1/Stj2/3/Ouasbar2	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
122	ADYT_120		ICARDA	Aghrass1/3/HFN94N8/Mrb5//Zna1/4/IcamorTA0458	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
127	IDON37-010		ICARDA	Marsyr3/3/Gcn//Stj/Mrb3	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
130	IDON37-021		ICARDA	ICAMORTA041/Quabrach1//Adnan1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
131	IDON37-022	Icacube	ICARDA	Mgnl3/Ainzen1//Mgnl3/Ainzen1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
132	IDON37-027		ICARDA	Mgnl3/Ainzen-1//Mgnl3/Aghrass2	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
133	IDON37-030		ICARDA	Mgnl3/Ainzen-1/3/Azeghar-1//Msbl-1/Quarmal	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
134	IDON37-033		ICARDA	Mgnl3/Ainzen-1//Ammar-1	yes	solid	9	ICARDA breeding program	9	ICARDA wide	5	Cultivars 2

135	IDON37-036	Margherita 2	ICARDA	Terbol975/Geruftel2	yes	solid	9	ICARDA breeding program	9	breeding programs exchange	5	Cultivars 2
136	IDON37-039		ICARDA	Mgnl3/Ainzen1/3/Ter1//Mrf1/Stj2	yes	solid	9	ICARDA breeding program	9	ICARDA wide	5	Cultivars 2
140	IDON37-052		ICARDA	Adnan2/Otb4//CM829/CandocrossH25	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
141	IDON37-053		ICARDA	IcamorTA042/4/Bcr/Lks4/3/Altar84/Stn//Lahn/5/Beltagy2/6/Ossl1/Stj5/5/Bicrederaa1/4/BezaizSHF//SD19539/Waha/3/Stj/Mrb3	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
143	IDON37-063		ICARDA	Ter1//Mrf1/Stj2/3/Icajihan22	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
145	IDON37-081		ICARDA	CM829/CandocrossH25/5/F413/3/Arthur71/Lahn//Blk2/Lahn/4/Quarmal	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
146	IDON37-082	Icadezful	ICARDA	Geromtel1/IRANYT053//Mgnl3/Ainzen1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
149	IDON37-095	Icarukus	ICARDA	Maamouri1/5/IcamorTA0462/4/Stj3//Bcr/Lks4/3/Icamor"s"/6/Mgnl3/Ainzen1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
150	IDON37-096		ICARDA	Azeghar2/5/IcamorTA0462/4/Stj3//Bcr/Lks4/3/Icamor"s"/6/Stj3//Bcr/Lks4/3/Ter3	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
151	IDON37-097		ICARDA	Mgnl3/Ainzen1/3/Bcr/Gro1//Mgnl1	yes	solid	9	ICARDA breeding program	9	breeding programs exchange	5	Cultivars 2
152	IDON37-105		ICARDA	Azeghar1//Blrn/Mrf2/3/Bicrederaa1/Azeghar2	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
155	IDON37-141		ICARDA	IcamorTA0471//IcamorTA0459/Ammar8/4/Stj3//Dra2/Bcr/3/Ter3	yes	solid	9	ICARDA breeding program	9	breeding programs exchange	5	Cultivars 2
156	IDON37-143		ICARDA	Mrb3/Mna1//Ter1/3/IcamorTA0459/Ammar7/4/Beltagy2	yes	solid	9	ICARDA breeding program	9	breeding programs exchange	5	Cultivars 2
158	IDYT37-03	Kunmiki	ICARDA	MorlF38//Bcrch1/Kund1149/3/Bicrederaa1/Miki	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
159	IDYT37-04	Berghouata1	ICARDA	Ter1//Mrf1/Stj2	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2

160	IDYT37-05	Icamoram7	ICARDA	ICAMORTA0472/Ammar7	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
162	IDYT37-08	Icakassem1	ICARDA	Geromtel1/Icasyr1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
164	IDYT37-10	Icarhani	ICARDA	Mrf1/Stj2//Gdr2/Mgnl1	yes	solid	9	ICARDA breeding program	9	breeding programs exchange	5	Cultivars 2
165	IDYT37-12	Icakassem2	ICARDA	Geromtel1/Icasyr1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
166	IDYT37-13	Icamoram8	ICARDA	Icamor/Ammar8	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
167	IDYT37-14	Maci115	ICARDA	Maamouri2/CI115/5/F413J.S/3/Arthur71/Lahn//Blk2/Lahn/4/Quar	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
168	IDYT37-15	Hessept	ICARDA	IcamorTA0462/4/Gdr2//(SwAlgia/Gdr1)43/3/IcamorTA0463/5/Ter1//Mrf1/Stj2	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
169	IDYT37-17	Icambel	ICARDA	Mrb3/Mna1//Ter1/3/ICAMORTA0459/Ammar7/4/Beltagy2	yes	solid	9	ICARDA breeding program	9	breeding programs exchange	5	Cultivars 2
170	IDYT37-18	Icarnada	ICARDA	Src2/Azn1/3/Bcr/Gro1//Mgnl1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
171	IDYT37-19		ICARDA	Mgnl3/Ainzen-1//Maamouri-3	yes	solid	9	ICARDA breeding program	9	breeding programs exchange	5	Cultivars 2
172	IDYT37-20	Miki3	ICARDA	Stj3//Bcr/Lks4	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
174	IDYT37-22	Second roue	ICARDA	Stj3//Bcr/Lks4/3/Ter3/4/Bcr/Gro1//Mgnl1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
177	DURUM_PANEL_UNIBO-061		ICARDA	BIC/3/CHAM1//GRA//STK	yes	solid	9	ICARDA breeding program	9	breeding programs exchange	5	Cultivars 2
179	DURUM_PANEL_UNIBO-063	CHACAN	ICARDA	Cham1/5/Cando/4/BY*2/Tace//II27655/3/TME//ZB/W*2	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
180	DURUM_PANEL_UNIBO-065		ICARDA	H.MOUL(MOR)/CHABA88_(HML/CHHB88)	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
182	DURUM_PANEL_UNIBO-068	Moulsabil2	ICARDA	H.mouline(Mor)/Sabil2	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
183	DURUM_PANEL_UNIBO-069	OMBAR	ICARDA	Mrb16/3/Ente/Mario//P66270/4/Ren/Bar	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
184	DURUM_PANEL_UNIBO-072		ICARDA	QUAD//ERP/MAL/3/UNKN	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
187	DURUM_PANEL_UNIBO-0118	AGHRASS1	ICARDA	Altar84/Stn//Lahn	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2

189	DURUM_PANEL_UNIBO-0121	AMEDAKUL1	ICARDA	PI321644/M8//Marazk/4/Bcr/3/Ch1//Gta/Stk	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
190	DURUM_PANEL_UNIBO-0122	AMMAR1	ICARDA	Lgt3/Bicrecham1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
191	DURUM_PANEL_UNIBO-0123	ARISLAHN5	ICARDA	Aristan/3/Lahn//Gs/Stk/4/Brch	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
193	DURUM_PANEL_UNIBO-0129	Bicrecham1	ICARDA	Bcr/3/Ch1//Gta/Stk	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
194	DURUM_PANEL_UNIBO-0130	BICRE	ICARDA	Bit/Creso	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
195	DURUM_PANEL_UNIBO-0131	Bicrederaa1	ICARDA	Bcr//Fg/Snipe/3/GdoVZ578/Swan//Dra2	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
199	DURUM_PANEL_UNIBO-0142	GUEROU1	ICARDA	Ato//Ibis/Fg	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
201	DURUM_PANEL_UNIBO-0146		ICARDA	Ouassel1/4/Buc/Chrc//Prl/3/Pvn/5/Hel/3/Bit/Corm//Shwa	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
204	DURUM_PANEL_UNIBO-0151	KABIR1	ICARDA	Ovi/Cp//Fg	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
205	DURUM_PANEL_UNIBO-0152	Gerboy	ICARDA	GdoVZ512/Cit//Ruff/Fg/3/Gr/Boy	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
207	DURUM_PANEL_UNIBO-0155	LAGONIL2	ICARDA	Lahn//Gs/Stk/3/Gil3	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
221	DURUM_PANEL_UNIBO-0186	Waldamez1	ICARDA	GdoVZ512/Cit//Ruff/Fg/3/DWL5023	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
222	MCHCB-079	Daki	ICARDA	Dack/Gediz//USDA575	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
227	MCHCB-0103		ICARDA	T.polonicumTurkeyIG45272/6/ICAMORTA0463/5/Mra1/4/Aus1/3/Scar/GdoVZ579//Bit	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
228	MCHCB-0110		ICARDA	Sebatel-1/Icasyr-1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
230	MCHCB-0126	Icasyr2	ICARDA	D68193A1A//Ruff/Fg/3/Mtl5/4/Lahn	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
231	MCHCB-0128	Geromtel3	ICARDA	Gsbl1/4/D68193A1A//Ruff/Fg/3/Mtl5	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
232	MCHCB-0133		ICARDA	Aghrass1//Msbl1/Quarmal	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
233	MCHCB-0154	Zna4	ICARDA	GdoVZ512/Cit//Ruff/Fg/3/Src3	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
241	MCHCB-0218		ICARDA	Azeghar2/4/Gnt/3/Gdfl/T.dicds20013//Bcr	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
124	ADYT_142	F11_00266 = Ouassara	ICARDA	Ouasloukos1/5/Azn1/4/BEZAIZSHF//SD19539/Waha/3/Gdr2	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2

128	IDON37-013	Icavicre	ICARDA	ICAMORTA0468/6/21563/AA//Fg/3/D68102A2A1A/4/Vitron/5/Bcr	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
129	IDON37-014	Ouassara1	ICARDA	Ouasloukos1/5/Azn1/4/BEZAIZSHF//SD19539/Waha/3/Gdr2	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
138	IDON37-049		ICARDA	IcaKader2	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
144	IDON37-071		ICARDA	Ter1/3/Stj3//Bcr/Lks4/4/Aghrass1/3/Mrf1//Mrb16/Ru	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
224	MCHCB-083		ICARDA	Cham5*4/Ae. speltoides 401294/4/ICAMOR-TA04-69/3/Bcr/Gro1//Mgnl1/5/Stj3//Bcr/Lks4/3/Ter-3	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
209	DURUM_PANEL_UNIBO-0159	MARSYR1	ICARDA	Msbl1/Quarmal	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
100	DAWRyT-0304		ICARDA	Korifla/AegSpeltoides Syr//Heider	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
103	DAWRyT-0317		ICARDA	Korifla/AegSpeltoides Syr//Mrb5	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
118	ADYT_046		ICARDA	IcamorTA041/4/IcamorTA0469/3/Bcr/Gro1//Mgnl1/5/MIKI2	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
119	ADYT_095		ICARDA	Ossl1/Stj5/5/Bicrederaa1/4/BEZAIZSHF//SD19539/Waha/3/Stj/Mrb3/6/Bcr/Gro1//Mgnl1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
125	IDON37-001	Icamator	ICARDA	IcamorTA041/4/Aghrass1/3/HFN94N8/Mrb5//Zna1/5/Malmuk1/Serrator1	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
126	IDON37-008	CaMdoH25	ICARDA	CM829/CandocrossH25	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
137	IDON37-048	Atlas1	ICARDA	Atl/4/Stn//Hui/Somo/3/Yav/Fg//Roh	yes	solid	9	ICARDA breeding program	9	ICARDA wide	5	Cultivars 2
148	IDON37-094		ICARDA	Aghrass1//Bezaiz982/Bcrch1/4/IcamorTA0462/3/Quabrach3//Vitron/Bidra1/5/Stj3//Bcr/Lks4/3/Ter3	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
208	DURUM_PANEL_UNIBO-0158	Maamouri1	ICARDA	1346/Lahn//Bcr/Lks4	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2

235	MCHCB-0161		ICARDA	319ADDO/5/D68193A1A//Ruff/Fg/3/Mtl5/4/Lahn	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
123	ADYT_134	F11_00257 = Ouassara	ICARDA	Ouasloukos1/5/Azn1/4/BEZAIZSHF//SD19539/Waha/3/Gdr2	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
220	DURUM_PANEL_UNIBO-0185	TUNSYR1	ICARDA	not available	yes	good	9	ICARDA breeding program	6	ICARDA wide	5	Cultivars 2
139	IDON37-051		ICARDA	Ossl1/Stj5/5/Bicrederaa1/4/BezaizSHF//SD19539/Waha/3/Stj/Mrb3/6/Icajihan11	yes	good	9	ICARDA breeding program	7	breeding programs exchange	5	Cultivars 2
142	IDON37-062		ICARDA	Ter1/3/Stj3//Bcr/Lks4/4/Icajihan18	yes	good	9	ICARDA breeding program	7	breeding programs exchange	5	Cultivars 2
154	IDON37-130	Bezajihan	ICARDA	Ossl1/Stj5/5/Bicrederaa1/4/BezaizSHF//SD19539/Waha/3/Stj/Mrb3/6/Icajihan12	yes	good	9	ICARDA breeding program	7	breeding programs exchange	5	Cultivars 2
382	DURUM_MAS	Waha	ICARDA	Plc/Ruff//Gta/Rtte	yes	good	9	ICARDA breeding program	7	breeding programs exchange	5	Cultivars 2
329	DURUM_PANEL_UNIBO-093	FLAMINIO	ITALY	LATINO/SENATORE-CAPPELLI	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
327	DURUM_PANEL_UNIBO-090	DON PEDRO	ITALY	Carc/Auk	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
328	DURUM_PANEL_UNIBO-091	DUILIO	ITALY	SENATORE-CAPELLI//ANHINGA/FLAMINGO,MEX	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
317	DURUM_PANEL_UNIBO-05	MERIDIANO	ITALY	SIMETO/WB-881//DUILIO/F-21	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
344	DWGIGSSALATFINA049	96367	MOROCCO	MAR85:112	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
345	DURUM_PANEL_UNIBO-047	NASSIRA	MOROCCO	INRAM.1805	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
349	DURUM_PANEL_UNIBO-056	OURGH	MOROCCO-TUNISIA	not available	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
352	DURUM_PANEL_UNIBO-059	YASMINE	MOROCCO-TUNISIA	not available	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
354	DURUM_PANEL_UNIBO-051	ANOUAR	MOROCCO-TUNISIA, INRA MAROC	not available	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2

			_Mar_1993									
358	DURUM_PANEL_UNIBO-0177	RAZZAK (TUN)	TUNISIA	21563/ANHINGA//FLAMINGO,MEX/3/DIEHL-MEDITERRANEAN/69.331[1281]; DIEHL-MEDITERRANEAN/69-331//KARIM	yes	solid	9	ICARDA breeding program	9	ICARDA	5	Cultivars 2
242	Jable-001	Yawa	AUSTRALIA	WT-LYLYT/R-875-LYT//LY-TM**[4172]**; WESTONIA/KALKA//KALKA/TAMAROI/3/RAC-875/KALKA//TAMAROI**[4172]**	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
245	Jable-004	Caparoi	AUSTRALIA	LY-2-6-3/930054	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
246	Jable-005	Bellaroi	AUSTRALIA	YALLAROI//TAM-1-B-17/KAMILAROI/3/TAM-1-B-17/KAMILAROI/3/DURATI(SIB)/LEEDS//GUILLEMOTS/4/WAGTAIL-10//SHEARWATER/MALLARD-DW	yes	good	10	CYMMIT breeding program	8	Dev countries	5	Cultivars 2
247	Jable-006	Jandaroi	AUSTRALIA	110780/111587; 920777/111566	yes	good	10	CYMMIT breeding program	8	Dev countries	5	Cultivars 2
250	Jable-009	Tamaroi	AUSTRALIA	ALTAR-84/4/TAM-1-B-17/KAMILAROI/3/WELLS/56111//GUILLEMOT	yes	good	10	CYMMIT breeding program	8	Dev countries	5	Cultivars 2
251	Jable-010	DBA Aurora	AUSTRALIA	HARD-FEDERATION/CLEVELAND//SANDS	yes	good	10	CYMMIT breeding program	8	Dev countries	5	Cultivars 2
243	Jable-002	WID802	AUSTRALIA	not available	yes	good	10	CYMMIT breeding program	8	Dev countries	5	Cultivars 2
244	Jable-003	Tjilkuri	AUSTRALIA	BRND-Y-DURAY-2/R-875-LYT//LY-TM[4172]; BRINDUR/3/YALLAROI*2//DUR-A/YALLAROI/4/RAC-875/KALKA//TAMAROI/5/LINGHZI/YALLA	yes	good	10	CYMMIT breeding program	8	Dev countries	5	Cultivars 2

				ROI//TAMAROI/3/LIN GHZI/YALLAROI								
249	Jable-008	Saintly	AUSTRAL IA	KALKA(SIB,WLYY-9)*2/TAMAROI	yes	good	10	CYMMIT breeding program	9	ICARDA	5	Cultivars 2
283	DURUM_PANEL_U NIBO-023		CIMMYT	CIMMYT198_(BUSH EN_4/TARRO_2//BU SHEN4)	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
285	DURUM_PANEL_U NIBO-025		CIMMYT	CIMMYT247_(RASC ON_37/2*TARRO_2)	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
292	DURUM_PANEL_U NIBO-0257		CIMMYT	1A.1D5+106/2*WB88 1//1A.1D5+106/3*MO JO/3/BISU_1/PATKA _3	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
296	DURUM_PANEL_U NIBO-0262		CIMMYT	GEDIZ/FGO//GTA/3/S RN_1/4/TOTUS/5/EN TE/MEXI_2//HUI/3/YA V_1/GEDIZ/6/SOMBR A_20/7/STOT//ALTAR 84/ALD	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
271	DURUM_PANEL_U NIBO-011		CIMMYT	CIMMYT23_(BISU_1/ PATKA_3)	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
274	DURUM_PANEL_U NIBO-014		CIMMYT	CIMMYT47_AHIOU1	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
275	DURUM_PANEL_U NIBO-015		CIMMYT	CIMMYT52_(KULRE NGIBALIKCIL_8)	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
276	DURUM_PANEL_U NIBO-016		CIMMYT	CIMMYT 67_PLATA 16	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
277	DURUM_PANEL_U NIBO-017		CIMMYT	CIMMYT73_PORTO5	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
278	DURUM_PANEL_U NIBO-018		CIMMYT	CIMMYT78_(ROK/FG O//STIL/3/BISU_1)	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
279	DURUM_PANEL_U NIBO-019		CIMMYT	CIMMYT104_YAZI10, 1	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
280	DURUM_PANEL_U NIBO-020		CIMMYT	CIMMYT108_(ACUAT ICO/YAZI_1)	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
281	DURUM_PANEL_U NIBO-021		CIMMYT	CIMMYT136_(FOCH A_1/5*ALAS)	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
282	DURUM_PANEL_U NIBO-022		CIMMYT	CIMMYT172_(TOPD Y_21/RASCON_33)	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
284	DURUM_PANEL_U NIBO-024		CIMMYT	CIMMYT222_(GS/CR A//SBA81/3/HO/MEXI _1/5/MEMO/6/2*…)	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
288	DURUM_PANEL_U NIBO-0253	Jupare C2003	CIMMYT	STOT//ALTAR84/ALD (=JUPAREC2003)	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
290	DURUM_PANEL_U NIBO-0255		CIMMYT	CHEN_1/TEZ/3/GUIL/ /CIT71/CII/4/SORA/P LATA_12/5/STOT//AL TAR84/ALD	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2

291	DURUM_PANEL_UNIBO-0256		CIMMYT	MALMUK_1//LOTUS_5/F3LOCAL(SEL.ETHIO.135.85)	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
287	DURUM_PANEL_UNIBO-0252	Atil C2000	CIMMYT	SOOTY_9/RASCON_37(=ATILC2000)	yes	good	10	CYMMIT breeding program	7	breeding programs exchange	5	Cultivars 2
293	DURUM_PANEL_UNIBO-0259		CIMMYT	AJAIA_12/F3LOCAL(SEL.ETHIO.135.85)//PLATA_13/3/SOMAT_3/4/SOOTY_9/RASCON_37	yes	good	10	CYMMIT breeding program	7	breeding programs exchange	5	Cultivars 2
299	DURUM_PANEL_UNIBO-0265		CIMMYT	PLATA_10/6/MQUE/4/USDA573//QFN/AA_7/3/ALBAD/5/AVO/HUI/7/PLATA_13/8/THKNEE_11/9/CHEN/ALTAR84/3/HUI/POC//BUB/RUFO/4/FNFOOT	yes	good	10	CYMMIT breeding program	7	breeding programs exchange	5	Cultivars 2
302	DURUM_PANEL_UNIBO-0269		CIMMYT	SOMAT_3/PHAX_1//TILO_1/LOTUS_4	yes	good	10	CYMMIT breeding program	7	breeding programs exchange	5	Cultivars 2
294	DURUM_PANEL_UNIBO-0260		CIMMYT	USDA595/3/D67.3/RABI//CRA/4/ALO/5/HUI/YAV_1/6/ARDENTE/7/HUI/YAV79/8/POD_9	yes	good	10	CYMMIT breeding program	9	ICARDA	5	Cultivars 2
297	DURUM_PANEL_UNIBO-0263		CIMMYT	VANRRIKSE_6.2//1A1D2+125/3*WB881	yes	good	10	CYMMIT breeding program	9	ICARDA	5	Cultivars 2
216	DURUM_PANEL_UNIBO-0179	SAJUR	ICARDA	Snipe/THunk	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
313	FIGSDRYWET134	107126	IRAN	IRN931:19	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
355	DURUM_PANEL_UNIBO-054	MARJANA	MOROCCO-TUNISIA, INRA MAROC_Mar_1996	D633/GAVIOTA/4/AVETORO//ANHINGA/PLE/3/DL672/SWAN//E	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
356	DURUM_PANEL_UNIBO-053	JAWHAR	MOROCCO-TUNISIA, INRA MAROC_Mar_1993	not available	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
370	DURUM_PANEL_UNIBO-035	BOLO	SPAIN	H-89092/H-89136//SOISSONS	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
371	DURUM_PANEL_UNIBO-036	BOMBASI	SPAIN	not available	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2

366	DURUM_PANEL_UNIBO-031	ASTIGI	SPAIN	not available	yes	good	10	CYMMIT breeding program	7	breeding programs exchange	5	Cultivars 2
375	DURUM_PANEL_UNIBO-041	GALLARETA	SPAIN-CIMMYT	RUFF/FLAMINGO,MEX//MEXICALI-75/3/SHEARWATER	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
378	DURUM_PANEL_UNIBO-043	JABATO	SPAIN-CIMMYT, ASGROW SEMILLAS_Esp_1996	not available	yes	solid	10	CYMMIT breeding program	10	CIMMYT deriv	5	Cultivars 2
19	DW-SR-FIGS119	96203	MOROCCO	MAR871:31	No							
20	FIGSDRYWET001	43315	OMAN	OMN87:113	No							
21	FIGSDRYWET003	79509	ETHIOPIA	ETH74::54	No							
24	FIGSDRYWET009	82029	KAZAKHSTAN	KAZS30	No							
65	FIGSDWHOTCLD015	81510	ETHIOPIA	ETH732::91	No							
67	FIGSDWHOTCLD051	84040	SAUDI ARABIA	SAUS14	No							
147	IDON37-085	Icaghram	ICARDA	IcamorTA0471//IcamorTA0459/Waha/3/Mgnl3/Ainzen1	No							
153	IDON37-129		ICARDA	CM829/CandocrossH25//Icajihan10	No							
157	IDYT37-01	Icarasha2	ICARDA	Stj3//Bcr/Lks4/3/Ter3	No							
161	IDYT37-06	Icajihan 2013	ICARDA	Mrf1/Stj2/3/1718/BT24//Karim	No							
163	IDYT37-09	Lahnmiki	ICARDA	Arislahn8//Bidra1/Miki	No							
175	IDYT37-23	Bezater	ICARDA	Ossl1/Stj5/5/Bicrederaa1/4/BEZAIZSHF//SD19539/Waha/3/Stj/Mrb3/6/Stj3//Bcr/Lks4/3/Ter3	No							
212	DURUM_PANEL_UNIBO-0164	NILE	ICARDA	Snipe/Fg	No							
337	DURUM_PANEL_UNIBO-0107	PLINIO	ITALY	LINEA-D-50/TRIGO-CANDEAL**[1620][1622**	No							

Annex 2: Heat susceptibility index (HSI) for GY, TKW, Gr.spk and Gr.m^{-2} calculated by comparing stressed environments (KED and Fan) to non-stressed environments (TES and MZ).

Trait	TKW15 (MZ-FAN)	TKW15 (MZ-KED)	TKW16 (MZ-FAN)	TKW16 (TES-FAN)	TKW16 (MZ-KED)	TKW16 (TES-KED)	HSI-TKW-(MOR-SR)	TKW BLUES	Gr.m^{-2} (MZ-FAN)	Gr.m^{-2} (TES-FAN)	Gr.m^{-2} (MZ-KED)	Gr.m^{-2} (TES-KED)	Gr.m^{-2} (MOR-SR)	Gr.m^{-2} BLUES
heat susceptible	104	115	123	119	118	116	117	103	125	112	121	109	100	113
TOTAL Heat tolerant	112	101	93	97	98	100	99	113	91	104	95	107	116	103
highly heat tolerant	17	10	38	78	48	51	67	28	69	2	94	4	4	89
heat tolerant	43	28	33	10	26	21	16	32	12	6	0	13	21	4
moderately heat tolerant	52	63	22	9	24	28	16	53	10	96	1	90	91	10

Trait	GY15 (MZ-FAN)	GY15 (MZ-KED)	GY16 (MZ-FAN)	GY16 (TES-FAN)	GY16 (MZ-KED)	GY16 (TES-KED)	HSI-GY-(MOR-SR)	GY BLUES	Gr.Spk (MZ-FAN)	Gr.Spk (TES-FAN)	Gr.Spk (MZ-KED)	Gr.Spk (TES-KED)	Gr.Spk (MOR-SR)	Gr.Spk BLUES
heat susceptible	107	104	109	113	108	107	105	88	111	106	109	109	104	102
TOTAL Heat tolerant	109	112	107	103	108	109	111	128	105	110	107	107	112	114
highly heat tolerant	35	17	32	2	34	2	0	8	86	11	90	11	19	78
heat tolerant	37	24	26	4	39	11	12	30	10	28	9	28	26	19
moderately heat tolerant	37	71	49	97	35	96	99	90	9	71	8	68	67	17

Annex 3: Significant QTLs with LOD and phenotypic variance (PV) for GY, TKW, SPK and PLH of the four populations.

Chr	QTL ID	Trait	Population	Marker	Position (cM)	LOD	PV (%)
1A	Qicd.SPK.001	SPK	IC	Icr_6218	111.3	5.3	8.0
1B	Qicd.TKW.001	TKW	IC, YG	Icr_6388	34.4	4.0	7.7
	Qicd.TKW.002	TKW	IC	c1B.loc125	124.4	6.1	15.9
2B	Qicd.GY.001	GY	YG	Icr_7246	69.3	5.0	13.4
	Qicd.GY.002	GY	IC	Icr_18720	107.0	4.4	6.2
	Qicd.TKW.GY.001	TKW	SW	c2B.loc125	123.7	4.7	8.6
		GY	YG	Icr_687	135.8	4.3	4.8
3B	Qicd.SPK.002	SPK	SW	Icr_5915	66.5	5.3	13.2
4A	Qicd.PLH.GY.001	PLH	DRO	Icr_21319	20.6	4.0	4.2
		GY	IC	c4A.loc33	32.9	3.8	5.1
	Qicd.PLH.GY.002	GY	SW	Icr_17289	43.8	6.2	19.3
		PLH	SW	Icr_1963	50.9	8.1	14.1
4B	Qicd.PLH.001	PLH	DRO	c4B.loc19	13.4	6.3	10.0
	Qicd.PLH.GY.003	GY	YG	Icr_987	32.9	7.5	9.4
		PLH	YG	Icr_7052	40.3	11.7	32.7
	Qicd.TKW.003	TKW	DRO, YG	Icr_16461	49.1	4.5	9.3
5B	Qicd.GY.003	GY	YG	Icr_15349	101.9	4.9	7.3
	Qicd.PLH.GY.004	PLH	DRO	c5B.loc116	112.9	4.0	3.9
		GY	DRO	c5B.loc121	116.9	4.1	6.0
	Qicd.PLH.002	PLH	YG	Icr_20534	154.1	6.6	8.4
	Qicd.SPK.003	SPK	DRO	Icr_2977	195.8	6.1	6.6
6A	Qicd.TKW.004	TKW	YG	Icr_15049	97.2	4.1	9.5
6B	Qicd.TKW.005	TKW	YG	c6B.loc6	5.1	6.1	4.7
	Qicd.PLH.003	PLH	YG	Icr_9314	79.0	4.5	15.8
	Qicd.SPK.004	SPK	SW	Icr_17674	173.1	6.1	10.9
7A	Qicd.TKW.006	TKW	DRO	Icr_5226	9.2	5.3	10.4
	Qicd.GY.004	GY	DRO	Icr_16322.	138.7	4.4	4.3
7B	Qicd.GY.005	GY	SW	Icr_14524	110.6	3.9	11.7

QTL ID: Qicd(icarda).trait.number, PV: Phenotypic variance (%), significant QTL $p < 0.05$

Annex 4: Frequency of re-identifying QTL associated to grain yield in the ten subset of training population.

Chr	QTL ID	Position (cM)	Frequency of QTL detection in the 10 TP subset			
			DRO	IG	YG	SW
2B	Qicd.GY.001	69.3	50%	-	30%	20%
2B	Qicd.GY.002	107.0	30%	40%	20%	-
2B	Qicd.TKW.GY.001	135.8	-	-	40%	-
4A	Qicd.PLH.GY.001	20.6	-	20%	-	20%
4A	Qicd.PLH.GY.002	43.8	-	-	-	30%
4B	Qicd.PLH.GY.003	32.9	10%	-	20%	-
5B	Qicd.GY.003	101.9	30%	10%	-	20%
7A	Qicd.GY.004	138.7	20%	20%	10%	20%
7B	Qicd.GY.005	110.6	-	20%	-	40%

frontiers in Plant Science

ORIGINAL RESEARCH
published: 18 July 2017
doi: 10.3389/fpls.2017.01277

Genetic Diversity within a Global Panel of Durum Wheat *(Triticum durum)* Landraces and Modern Germplasm Reveals the History of Alleles Exchange

*Hafssa Kabbaj[1,2], Amadou T. Sall[1,2], Ayed Al-Abdallat[3], Mulatu Geleta[4], Ahmed Amri[1], Abdelkarim Filali-Maltouf[2], Bouchra Belkadi[2], Rodomiro Ortiz[4] and Filippo M. Bassi[1]**

[1] International Center for Agricultural Research in the Dry Areas, Rabat, Morocco, [2] Department of Plant Science, Mohammed V University, Rabat, Morocco, [3] Department of Horticulture and Crop Science, Faculty of Agriculture, The University of Jordan Amman, Amman, Jordan, [4] Plant Breeding, Swedish University of Agricultural Sciences, Alnarp, Sweden

OPEN ACCESS

Edited by:
Agata Gadaleta,
Università degli Studi di Bari Aldo Moro, Italy

Reviewed by:
Steven Shujun Xu,
United States Department of Agriculture, United States
Daniela Marone,
Centre of Cereal Research, CREA-CER, Italy

***Correspondence:**
Filippo M. Bassi
f.bassi@cgiar.org

Specialty section:
This article was submitted to Crop Science and Horticulture, a section of the journal Frontiers in Plant Science

Received: 25 May 2017
Accepted: 06 July 2017
Published: 18 July 2017

Citation:
Kabbaj H, Sall AT, Al-Abdallat A, Geleta M, Amri A, Filali-Maltouf A, Belkadi B, Ortiz R and Bassi FM (2017) Genetic Diversity within a Global Panel of Durum Wheat (Triticum durum) Landraces and Modern Germplasm Reveals the History of Alleles Exchange. Front. Plant Sci. 8:1277. doi: 10.3389/fpls.2017.01277

Durum wheat is the 10th most important crop in the world, and its use traces back to the origin of agriculture. Unfortunately, in the last century only part of the genetic diversity available for this species has been captured in modern varieties through breeding. Here, the population structure and genetic diversity shared among elites and landraces collected from 32 countries was investigated. A total of 370 entries were genotyped with Axiom 35K array to identify 8,173 segregating single nucleotide polymorphisms (SNPs). Of these, 500 were selected as highly informative with a PIC value above 0.32 and used to test population structure via DAPC, STRUCTURE, and neighbor joining tree. A total of 10 sub-populations could be identified, six constituted by modern germplasm and four by landraces of different geographical origin. Interestingly, genomic comparison among groups indicated that Middle East and Ethiopia had the lowest level of allelic diversity, while breeding programs and landraces collected outside these regions were the richest in rare alleles. Further, phylogenetic analysis among landraces indicated that Ethiopia might represent a second center of origin of durum wheat, rather than a second domestication site as previously believed. Together, the analyses carried here provide a global picture of the available genetic diversity for this crop and shall guide its targeted use by breeders.

Keywords: center of diversity, couscous, domestication, evolution, pasta, Axiom 35K, array, durum wheat

INTRODUCTION

Durum wheat (*Triticum turgidum* ssp. *durum* Desf., $2n = 4x = 28$, AABB) is the 10th most important crop worldwide owing to its annual production of 37 million tons (LMC International, 2009; Ranieri, 2015; Taylor and Koo, 2015). It is grown on about 10% of the world's wheat area mostly in West Asia, North, and East Africa, the North American Great Plains, India, Eastern and Mediterranean Europe (Cantrell, 1987; International Wheat Council, 1991). With the exception of Europe, North Africa (Algeria, Morocco, Tunisia, and Libya) is the largest import market for durum wheat (Bonjean et al., 2016). Its final uses vary between industrial production of pasta,

couscous, and other semolina products and traditional handmade foods such as *frike*, bourghul, and unleavened breads. The vast array of homemade foods derived from durum grains is the result of its long history as part of human diets, which dates back to the origin of civilization in the Fertile Crescent (MacKey, 2005). Tetraploid wheat domestication took place about 12,000 years ago in the Fertile Crescent, when ancient farmers selected among cultivated forms of wild emmer (*Triticum turgidum* ssp. *dicoccoides*) a naked type that was easier to thresh (*Triticum turgidum* ssp. *dicoccum*; MacKey, 2005; Tanno and Willcox, 2006; Zohary et al., 2012). Approximately 2,000 years after this event, human migration and the spread of agriculture from the Fertile Crescent to and throughout Europe and Asia led to the expansion of the cultivation of naked emmer. During the same period, durum wheat (*Triticum turgidum* ssp. *durum*) appeared in the Fertile Crescent as result of further selection and domestication of naked emmer (Zohary et al., 2012). Due to its larger grains and higher productivity, durum gradually replaced its ancestor to become by the second millennium BC the major cultivated form of tetraploid wheat (Maier, 1996; Nesbitt and Samuel, 1998; Zohary et al., 2012).

Thus, durum wheat origin is the result of two successful domestication events by ancient farmers, first from wild emmer to domesticated emmer, and second from cultivated naked forms of emmer to durum (Gioia et al., 2015). The Levantine (Jordan, Lebanon, Israel, Palestine, and Syria) is considered to be the center of origin of this crop (Vavilov, 1951; Feldman, 2001). From there, it spread throughout the Mediterranean basin, probably via trading by Phoenician merchants, by the caravans' routes along the Sahara desert or the North African coasts (Bozzini, 1988), and the Silk Road to Asia (Waugh, 2010). Reports (Mengistu et al., 2015, 2016) suggested, that durum wheat was also domesticated a third time to derive *Triticum aethiopicum* Jakubz. (syn. *Tritium durum* subsp. *abyssinicum* Vavilov), which is mainly found today under cultivation in Ethiopia and neighboring countries. It remains yet unclear if this additional domestication was the result of further modification by farmers' of a durum landrace population originated in the Levantine, or rather if it represented a novel origin of durum by a separate domestication of naked emmer. What is clear is that the *abyssinicum* subspecies is morphologically very different, with uncompact spikes and small dark seeds (Sakamoto and Fukui, 1972; Porceddu et al., 1973; Pecetti et al., 1992; Mengistu et al., 2015).

The history of the durum wheat genetic makeup became more complex at the beginning of the 20th century when breeders started imposing artificial hybridization and selection pressure for commercial purposes (Autrique et al., 1996; Pecetti and Annicchiarico, 1998). In 1910, Nazareno Strampelli set up the first durum wheat breeding program in Foggia, southern Italy. This program was initially based on the selection of pure lines from local landraces (Scarascia Mugnozza, 2005). Later, Strampelli recognized the great value of the inheritance laws described by Mendel and started a true hybridization program. The most successful result was the cultivar 'Cappelli,' released in 1915 (Laidò et al., 2013). This pioneer cultivar had a major global impact in the years that followed, and most of the modern varieties can be traced back to the 'Capelli' lineage.

A second major impact was provided by the shuttle breeding system developed by the Nobel laureate Norman E. Borlaug several years later in Mexico and his deployment of dwarfing genes to increase harvest index (Gale and Youssefian, 1985). This resulted in the release of several semi-dwarf and widely adapted cultivars that are still grown nowadays (Ortiz et al., 2007). The modern scenario of the pedigrees post Green Revolution is extremely hard to describe, with several hybridization occurring between different breeding programs and mega-cultivars that have crossed the boundaries of their country of origin. To disentangle the last 40 years of germplasm exchange and cross hybridizations, new methods have been devised based on the allelic similarities described by molecular markers (Pritchard and Rosenberg, 1999; Christiansen et al., 2002; Falush et al., 2003; Flint-Garcia et al., 2003). A bi-product of these type of studies is the understanding of how much of the overall available alleles (namely: genetic diversity) have been captured within a specific germplasm. Since genetic diversity is often seen as an essential source of novel and useful alleles to be selected by breeders (Tanksley and McCouch, 1997; Cooper et al., 2001; Spillane and Gepts, 2001; Acosta-Gallegos et al., 2007), these types of studies have both a historical value and an immediate practical impact on breeding. Hence, the aim of this research was to conduct a molecular assessment of a global durum wheat collection of cultivars, elite breeding lines and landraces, in order to photograph the current state of germplasm exchange and overall available genetic diversity.

MATERIALS AND METHODS

Plant Material

A large durum wheat germplasm collection exceeding 1,500 accessions was assembled at the field station of the International Center for Agricultural Research in the Dry Areas (ICARDA) in Terbol, Lebanon (33° 49′ 05′′ N, 35° 58′ 59′′ E). A core subset was defined after assessing the collection for similarity in flowering time, response to toxic level of boron, disease response, tendency to lodge, visual selection, and characterized with 10 single nucleotide polymorphisms (SNPs) associated to known genes. The original set contained several landraces selected on the basis of the algorithm for Focus Identification of Germplasm Sources (FIGS; Mackay et al., 2005; Bari et al., 2012) targeting the model to identify sources of rust resistance, tolerance to drought, heat and mineral toxicity. A core subset of 384 accessions was selected to be similar in phenology and diverse for all other traits. It includes 96 landraces from 24 countries and 288 cultivars and elite breeding lines from eight countries, ICARDA, and International Maize and Wheat Improvement Center (CIMMYT). This panel was built on the work already carried on by Maccaferri et al. (2003), removing duplicates and adding a set of landraces, new breeding material from ICARDA, CIMMYT, Canada, and Australia (**Supplementary Figure S1**). A detailed list of materials is provided in **Supplementary Table S1**.

DNA Extraction and Genotyping

DNA was extracted from leaf samples using a standard cetyltrimethylammonium bromide (CTAB) protocol (Doyle and Doyle, 1987). The 384 accessions were genotyped by 35K Affymetrix Axiom wheat breeders array[1] at Trait Genetics (Gatersleben, Germany) following the manufacturer instructions. This array was developed by choosing tags of proven high polymorphism when tested on modern bread wheat elites, among the 817k SNP Axiom HD platform.

Data Analysis

The polymorphic information content (PIC) was calculated following the formula described by Botstein et al. (1980), and the 2-points LOD was generated using Carthagene software option 'SEM' (Schiex et al., 2009). The discriminant analysis of principal components (DAPC), was performed using the 'adegenet' package 1.4-1 (Jombart et al., 2010) in R studio V 2.3.2 (R Development Core Team, 2011). With DAPC, the genetic variation was decomposed using a multivariate ANOVA model as:

$$\text{Total variance VAR}(X) = \text{variance between groups B}(X) + \text{variance within groups W}(X)$$

Other approaches such as principal component analysis (PCA) or principal coordinates analysis (PCoA/MDS) focus on VAR(X). That is, they only describe the global diversity, possibly overlooking differences between groups (Jombart and Collins, 2015). The variance explained by PCA was fixed to 75 and the value of k was tested from 2 to 50. The rate of decrease of the Bayesian information criterion (BIC) was visually examined (**Figure 1**), and the number of clusters was determined as the value of k above which BIC values decreased. Analysis of admixture by kinship was performed using the Bayesian clustering algorithm implemented in the software STRUCTURE v 2.3.4 (Pritchard et al., 2000) using 50,000 burning periods and 10,000 replicates and re-assessed five times with 11 independent runs. The value of k was set based on DAPC results. To further confirm cluster analysis, unweighted pair group method of association (UPGMA) was carried out using the genetic similarity matrix by numerical taxonomy and multivariate analysis system (NTSYS-PC) version 2.02e software (Rohlf, 1997). Because this method uses genetic similarity matrix, a line of reference was arbitrary set to explain 60% of similarity in order to determine the genetically distinct branches of the tree. Arlequin 3.5.2.2 (Excoffier et al., 2005) was used to assess the molecular variance (AMOVA) between clusters. Phylogenetic studies of landraces were conducted by neighbor-joining algorithm of the genetic distances determined by STRUCTURE using 1,000 bootstrapping analysis for an unrooted tree by DARwin V 6.0.12 software (Perrier et al., 2003). DIVA-GIS V 7.5.0 software (Hijmans et al., 2012) was used to graphically map the GPS coordinates of the places of collection of the landraces.

[1] www.affymetrix.com

RESULTS

Genotyping of a Global Panel of Durum Elites, Cultivars, and Landraces

A total of 384 durum entries were genotyped, but only 370 showed DNA quality sufficient for SNP calls. In total, 35,143 SNPs were assessed, of these 11,642 (34%) failed to meet the minimum call rate, which suggests that these markers were probably located on the D genome, present in hexaploid bread wheat but not in tetraploid durum wheat. A total of 14,851 (42%) met the quality cutoff but remained monomorphic in this population, while 8,173 (36%) were found to be high quality and polymorphic. The average frequency of the minor allele was 12% with a minimum of 3%. **Table 1** provides the chromosome assignment of markers based on the work by Winfield et al. (2016) in bread wheat, where 1,559 markers remained unassigned.

Population Stratification

A subset of 500 highly polymorphic ($0.32 \leq \text{PIC} \geq 0.45$) markers was chosen for clustering and kinship studies. These markers were selected for even distribution across the genome, covering

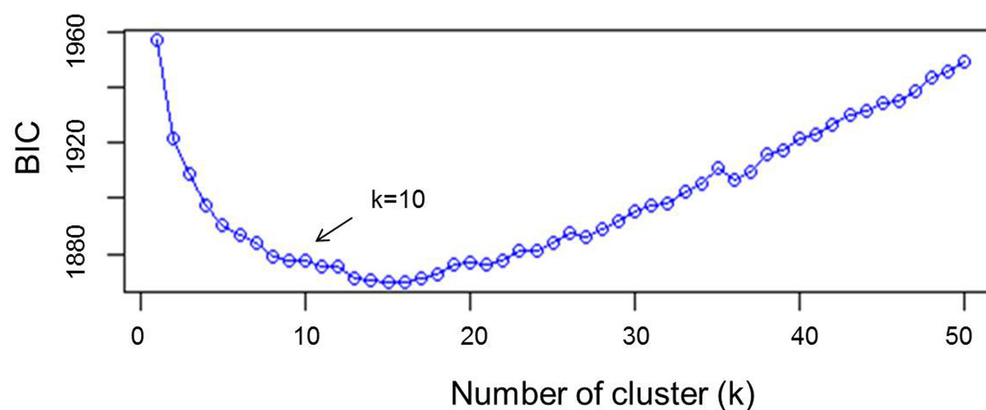

FIGURE 1 | Statistical determination of the optimum number of clusters by discriminant analysis of principal components (DAPC).

TABLE 1 | Number and distribution across the 14 chromosomes (Chr.) of durum wheat of polymorphic SNPs markers on the Axiom 35K breeder's array and the 2-points LOD for the subset of the 500 SNPs used for clustering.

Chr.[a]	Polymorphic	Subset of 500 SNPs	2-points LOD of the subset		
			Average	MIN	MAX
1A	505	26	36.7	0.3	101.2
1B	617	48	33.9	0.0	111.1
2A	519	31	36.8	0.0	111.1
2B	589	28	23.7	0.0	110.8
3A	411	22	36.8	0.0	110.5
3B	533	41	30.8	0.1	111.4
4A	306	25	35.8	0.0	101.5
4B	283	22	39.6	0.0	110.8
5A	489	33	35.7	0.0	110.2
5B	673	48	31.4	0.0	110.2
6A	360	22	49.4	0.0	111.1
6B	480	40	26.1	0.0	111.1
7A	505	35	36.5	0.0	110.5
7B	344	41	41.4	0.0	111.4
Unassigned	1,559	38	8.5	0.0	78.0
Total	8,173	500			

[a] *Chromosome assignment was done on the basis of a consensus bread wheat genetic map (Winfield et al., 2016).*

all durum chromosomes, with LOD values that ranged from 0 to 111.1, and averaged at a minimum of 23.7 in chromosome 2B (**Table 1**). These LOD scores indicate good distribution and correct chromosome assignment. DAPC inferred the optimum number of sub-populations to be 10 (**Figure 1**). AMOVA was used to determine that variation among and within groups was highly significant ($P < 0.001$), with the clusters capturing 31.5% of the total genetic variations, while 68.3% was explained by individuals within populations (**Table 2**). Among the 10 clusters, four groups were composed of landraces, while six groups included mostly cultivars and elite lines (**Figure 2**). STRUCTURE was also used to determine cluster assignment, with the strongest contradiction between DAPC and STRUCTURE identified among landraces. In fact, only two clusters were identified by the latter compared to the four by DAPC (**Figure 3**). This issue was circumvented by running separately landraces and modern lines in STRUCTURE, in which case good agreement could be found between the two software. Instead, tree-based studies by UPGMA identified six clusters determined at 60% of similarities, three composed of mostly landraces and three by modern lines. This value was kept as identified. Entries were assigned to cluster based on DAPC study as it was considered the most reliable method, but a qualitative score was given to each assignment as 'solid' when the three methods agreed in the assignment, 'good' when two methods agreed, and 'bad' when only DAPC made the assignment. Among 370 lines, 16 (4.3%) were scored as 'bad,' 70 (18.9%) as 'good,' and 284 (76.7%) as 'solid' (**Table 3**). In particular, sub-populations 3 and 6 had the highest number of 'bad' assigned entries. Clusters 1, 2, 5, 8, and 9 were the most solid with no 'bad' assignments, and few 'good.' Full details for each genotype are provided in **Supplementary Table S1**.

TABLE 2 | Analysis of molecular variance (AMOVA) for stratification of a global durum wheat panel into 10 sub-populations.

SOV	d.f.	Sum of squares	Variation (%)
Populations (P)	9	20,906***	0.315
Individuals within P	360	49,809***	0.683
Individuals	370	67***	0.002
Total	739	70783	

****Indicate that source of variation was highly significant at $P \leq 0.001$.*

Cluster 1 comprises 11 landraces from West Asia (Levantine). Cluster 2 is represented by 18 landraces, 15 from Ethiopia, 1 from Yemen, 1 from Jordan, and 1 from Russia. Cluster 3 is composed of 26 landraces, 5 from Tunisia, 4 from Algeria and Spain, 2 from Afghanistan, Greece, and Italy, and 1 each from Azerbaijan, China, Ethiopia, Iran, Kazakhstan, and Russia. Cluster 4 is composed mainly of Central and South Asian landraces and includes those collected in Afghanistan, Armenia, Georgia, India, Iran, Iraq, Kazakhstan, Pakistan, Turkey, Russia, but also Italy, Oman, Yemen, and Saudi Arabia. Cluster 5 gathers 13 modern lines from the breeding program of ICARDA, which include in their pedigree 'Om Rabi' – a line derived from the cross between the elite 'Jori' and the Jordanian landrace 'Haurani,' one Italian landrace and the Italian cultivars 'Arcangelo,' 'Appio,' and 'Capeiti.' Cluster 6 contains 20 modern lines and 6 landraces, with 13 cultivars and 2 landraces from Italy, 4 landraces from Ethiopia, and the remaining modern germplasm from France, ICARDA and Spain. Cluster 7 is represented by 58 entries from different breeding programs, including 24 elites from ICARDA, 6 from CIMMYT, 4 cultivars from France and USA, 7 from

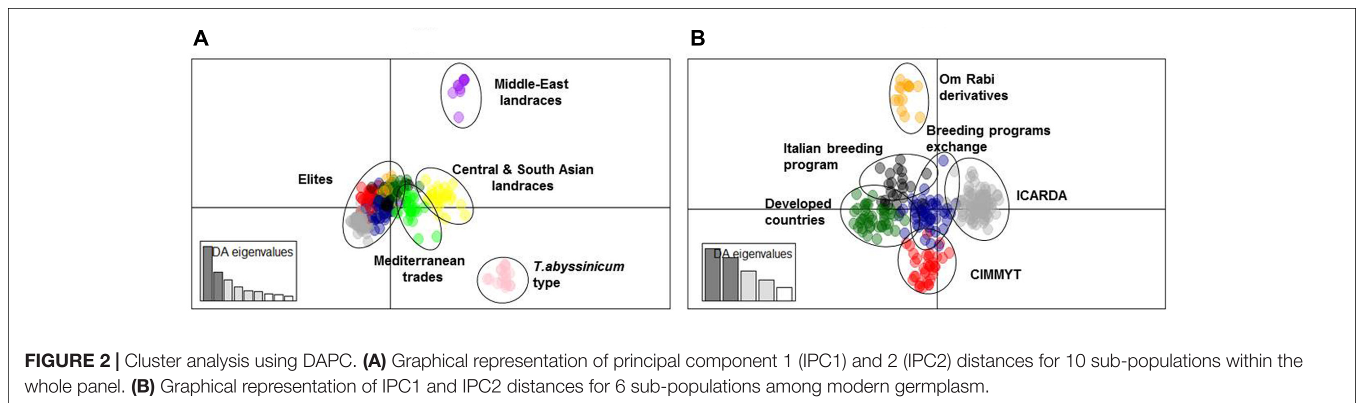

FIGURE 2 | Cluster analysis using DAPC. **(A)** Graphical representation of principal component 1 (IPC1) and 2 (IPC2) distances for 10 sub-populations within the whole panel. **(B)** Graphical representation of IPC1 and IPC2 distances for 6 sub-populations among modern germplasm.

FIGURE 3 | Admixture analysis by kinship of 370 durum wheat landraces and modern elites. Each individual is represented by a horizontal line. Color codes follow the color assignment of **Figure 2**, with the exception of clusters 1 to 4 that were merged into one color (one cluster) when tested by STRUCTURE (light blue).

Italy, 6 from Morocco and Spain, but also 1 landrace each from Spain and France. Cluster 8 includes 17 cultivars from North America, 5 from Australia, 2 from France, 1 from Italy and Spain, plus 2 landraces from Algeria and 1 landrace selection ('Shabha') from ICARDA. Cluster 9 is the largest with 106 breeding lines from ICARDA, 2 from CIMMYT, 4 varieties from Italy, 4 from Morocco and Tunisia, 'Wallaroi' from Australia, and 2 Moroccan landraces. Cluster 10 groups 24 elite lines derived from the breeding program of CIMMYT, 1 from ICARDA, 9 Australian cultivars, 5 Spanish, 2 Moroccan, and 1 Iranian landrace. The clustering of the panel is presented in **Figure 2**, and modern lines are detailed in **Figure 2B**.

TABLE 3 | Reliability of the entries assignment based on comparison between three genetic clustering methods: DAPC, STRUCTURE, and neighbor-joining (details available in **Supplementary Table S1**).

DAPC Cluster IDs	Assigned entries (*N*)	Reliability score		
		Bad	Good	Solid
1. Middle East	11	0	0	11
2. *T. abyssinicum* type	18	0	0	18
3. Mediterranean trades	26	7	8	11
4. Central and South Asian	27	1	4	22
5. 'Om Rabi' derivatives	13	0	1	12
6. Italian cultivars	26	6	5	15
7. Breeding program exchange	58	2	29	27
8. Developed countries	30	0	4	26
9. ICARDA derived	119	0	5	114
10. CIMMYT derived	42	0	14	28
Total	370	16	70	284

Admixture Analysis by Kinship

Admixture analysis was conducted using STRUCTURE. To better detail the kinship among landraces, these were also analyzed alone in the form of a phylogenetic analysis (**Figure 4**). Four main branches could be identified as defined by clusters 1, 2, 3, and a part of 4. In fact, cluster 4 containing mostly Central and South Asian landraces had the highest level of admixture, it located toward the origin of the tree, and created one independent branch with seven landraces from Armenia, Kazakhstan, India, Russia, Afghanistan, Turkey, and Georgia. Still, landraces from the same cluster also distributed to the branches generated by cluster 2 and 3. Cluster 2 containing mostly landraces from Ethiopia generated the branch further away from the origin of the tree, with landraces of cluster 4 from Iraq, Afghanistan, India, Pakistan, Yemen, Saudi Arabia, and Oman that located along this branch (**Figure 4**). Cluster 1 generated an independent phylogenetic path, with landraces from Syria, Jordan, and Iraq composing the edge of the branch. Cluster 3 included landraces from many countries. The coordinates of the collection sites of the landraces were placed on a map and color coded to match their cluster assignment (**Figure 5**). The admixture level among elites was higher than for the landraces (**Figure 3**). Cluster 9 containing ICARDA breeding lines was the largest group, and it is therefore unsurprising that it also presents the highest level of admixture, followed by cluster 7 of *breeding exchange*. Cluster 6 of the *Italian breeding program* has noticeable allelic similarity with cluster 3 of *Mediterranean trades*, but also with cluster 8 of breeding programs from *developed countries*. The cluster of *Om Rabi derivatives* is the smallest group, which shows a low level of admixture, but still it has some alleles in common with the cluster of ICARDA breeding program. Also the CIMMYT's elites group shared an important amount of alleles with the latest cluster.

Genetic Diversity among 10 Sub-populations

A second set of 500 markers capturing the rarest alleles ($0.005 \leq$ PIC ≥ 0.01) and the full set of 8.2K polymorphic markers were used to further assess allelic frequencies (**Table 4**). Cluster 1 presented the lowest genetic diversity with PIC = 0.04, followed by Cluster 5 (PIC = 0.05), these two clusters also had the lowest frequency of segregating markers 0.13 and 0.19, respectively, and did not capture any of the rare alleles.

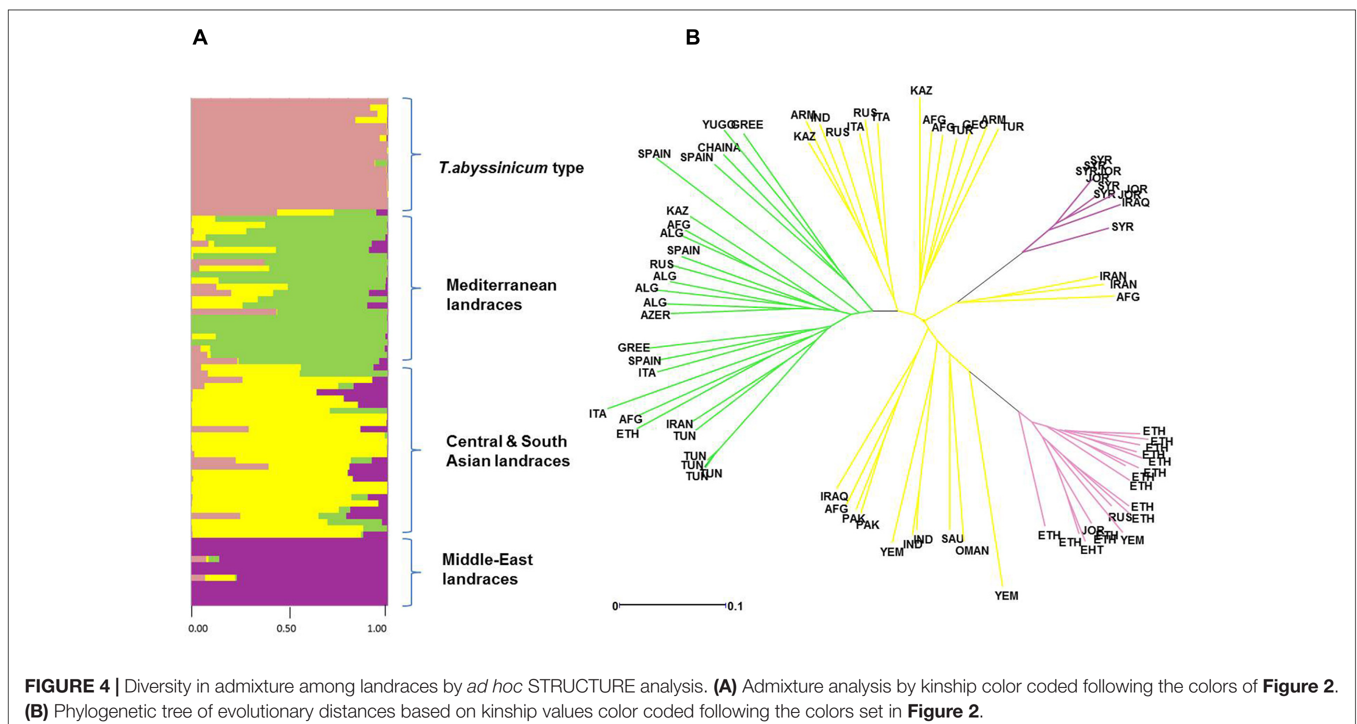

FIGURE 4 | Diversity in admixture among landraces by *ad hoc* STRUCTURE analysis. **(A)** Admixture analysis by kinship color coded following the colors of **Figure 2**. **(B)** Phylogenetic tree of evolutionary distances based on kinship values color coded following the colors set in **Figure 2**.

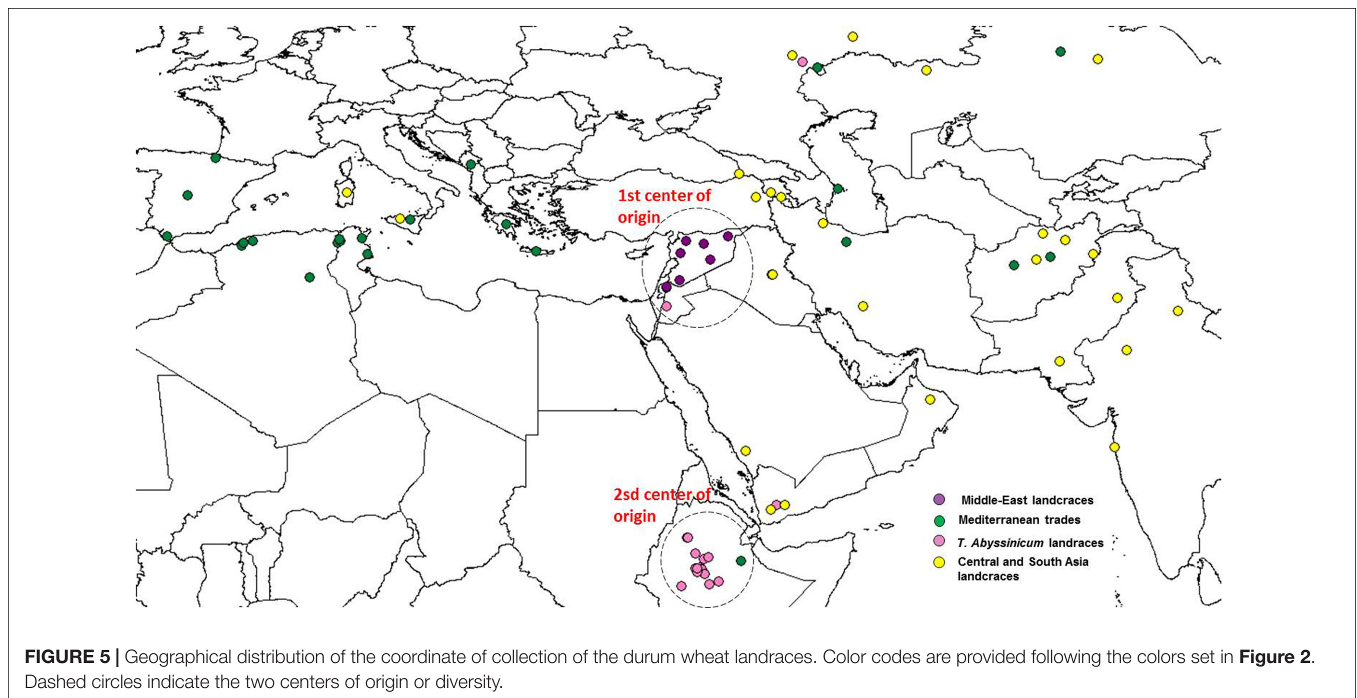

FIGURE 5 | Geographical distribution of the coordinate of collection of the durum wheat landraces. Color codes are provided following the colors set in **Figure 2**. Dashed circles indicate the two centers of origin or diversity.

TABLE 4 | Genetic diversity captured by each sub-population.

		All markers			Rare alleles		
DAPC cluster IDs	***N***	**Fixed[a] (%)**	**Segregating (%)**	**PIC[b]**	**Fixed[a] (%)**	**Segregating (%)**	**PIC**
1. Middle East	11	0.77	0.13	0.04	1.00	0.00	0.005
2. *T. abyssinicum* type	18	0.47	0.42	0.13	1.00	0.01	0.005
3. Mediterranean trades	26	0.35	0.65	0.17	0.94	0.06	0.005
4. Central and South Asian	27	0.15	0.84	0.22	0.38	0.62	0.005
5. 'Om Rabi' derivatives	13	0.78	0.19	0.05	1.00	0.00	0.005
6. Italian cultivars	26	0.53	0.46	0.11	0.99	0.01	0.005
7. Breeding program exchange	58	0.42	0.58	0.14	0.96	0.04	0.005
8. Developed countries	30	0.56	0.43	0.12	0.99	0.01	0.005
9. ICARDA derived	119	0.48	0.51	0.11	0.97	0.03	0.005
10. CIMMYT derived	42	0.56	0.44	0.10	0.79	0.21	0.005

[a] *Refers to major allele.*
[b] *Polymorphism information content.*

The group of Ethiopian landraces (cluster 2) also had a low level of segregating alleles (42%) comparable to what was observed for breeding programs from developed countries (cluster 8), and equally captured very few rare alleles (1%). The Central and South Asia landraces was the most genetically diverse sub-population, with a PIC of 0.22, almost all markers segregating (0.84) and also captured 62% of the rare alleles. This was also the cluster with the highest admixture that distributed along all other phylogenetic branches. Among the modern germplasm, clusters 7 and 9 were the most diverse, with 58 and 51% of total alleles captured, and 4 and 3% of rare alleles represented in the sub-populations, respectively. Interestingly, cluster 10 maintained the highest level of rare alleles (21%) among the modern germplasm.

DISCUSSION

Genetic diversity is of paramount importance as a source of novel traits and alleles for plant breeding, particularly to face the unpredictable challenges laying ahead, at a time of changing climates and new end-user demands (Tester and Langridge, 2010). However, diversity *per se* is of limited use (Frankel et al., 1995; Royo et al., 2009; Novoselović et al., 2016). It is instead to the breeders' advantage to know which ideal sources of diversity should be integrated within each program to better target their crossing schemes. With this scope, the global diversity of durum wheat was assessed comparing breeding efforts, historical cultivars, and landraces from 28 countries. Genotyping with the Axiom 35K "breeders' array" revealed that the panel used

could capture 36% of total polymorphism existing for the A and B genomes markers available on the array. A subset of 500 highly informative SNP markers was used to assess the genetic structure and stratification of the panel. This number of markers was in excess as compared to what previously reported in the literature (Maccaferri et al., 2003; Raman et al., 2010; Royo et al., 2010; Cabrera et al., 2015). Overall, the genotyping results were satisfactory and allowed the implementation of all downstream applications.

Success Level of the Clustering Procedure

Human practices such as farming, consumption habits, and trading of seeds within and among communities generate pressure bias, drift or founder effects on the germplasm (Dyer and Taylor, 2008; Delêtre et al., 2011). Furthermore, societal, cultural, and natural barriers reinforce reproductive isolation, limiting or encouraging gene flow among cultivars (Pusadee et al., 2009; Deu et al., 2013). Thus, several factors can influence the genetic diversity within a germplasm collection and the analysis presented here can only explain a fraction of it. The results of the AMOVA confirmed that the DAPC model was able to capture approximately one fifth of the total variance by stratifying the panel into 10 clusters, with individuals that maintained high levels of genetic diversity within groups. Thus, even if the choice of $k = 10$ was conservative as shown by the AMOVA, it was considered adequate to better identify similarities between genotypes, rather than over-fit their differences (Jombart et al., 2010). Clustering landraces by allelic similarities (kinship and admixture) is *de facto* an attempt of tracing those alleles that are identical by descent, hence maintained from their original domestication event or shared environmental pressures. Instead, in the case of cultivars and elite lines, genetic similarity is strongly influenced by the breeders' subjective choice of hybridizing specific germplasm sources to develop new lines. Since different breeding programs tend to utilize the same founders in their crossing strategies, strong admixtures exist between geographically distant germplasm. Separating cultivars and elite lines into groups of shared allelic similarity is therefore an attempt to capture the complex hybridization history of the breeding germplasm.

The population stratification was done via DAPC (**Supplementary Figure S2**), STRUCTURE analysis, and neighbor joining method. While some disagreement was found among methods, only 4.3% of tested genotypes were not assigned to the same clusters by two or more of the software used for the analysis (**Table 3** and **Supplementary Table S1**). Considering the high level of admixture showed by the landrace sub-group identified by STRUCTURE, and the clustering by UPGMA tree (**Supplementary Figure S2**), the results of DAPC were deemed more reliable.

History of Durum Origin and Migration Based on Genetic Diversity of Landraces

The population stratification of this panel identified four groups of landraces and six groups of cultivars and elite lines. Cluster 1 includes landraces from Jordan, Syria, and Iraq (**Figure 4**), countries that correspond to the center of origin of durum wheat in the Levantine (MacKey, 2005). The geographical proximity of these landraces to the center of origin maintained a high level of genetic purity with low levels of admixture (**Figure 4**) and almost complete fixing of major (77%) and minor (90%) alleles (**Table 4**). This is in good agreement with what was reported previously for landraces from Jordan (Rawashdeh et al., 2007; Mohammadi et al., 2014). More interestingly, the phylogenetic tree (**Figure 4**) clearly indicates how the germplasm from Syria and Jordan are more closely related as compared to the germplasm from Iran and Afghanistan belonging to cluster 4. This would suggest that durum wheat truly originated in the South end of the Fertile Crescent (MacKey, 2005), and only later migrated to the neighboring regions.

Cluster 2 groups mostly landraces from Ethiopia, with the exception of one from Yemen, one from Jordan, and another from Russia. However, the Russian landrace appears as wrongly assigned based on its high level of admixture (**Figure 4**). Ethiopia is known as a "secondary center of durum wheat diversity" (Harlan, 1969; Vavilov, 1992). Landraces from this country have unique morphology (Sakamoto and Fukui, 1972; Porceddu et al., 1973; Pecetti et al., 1992) and represent a separate sub-species under the name *T. durum* subs. *abyssinicum* or *T. aethiopicum* (Mengistu et al., 2015, 2016). **Figure 4** clearly shows that this germplasm is distinct from the primary region of origin of durum wheat (Middle-East landraces) with substantially no kinship to it. Furthermore, there is limited admixture between this group and any others. Hence, Ethiopia truly represents a center of diversity for durum wheat, without an evident allelic similarity to the primary origin in the Levantine, as also suggested by several other authors (Sakamoto and Fukui, 1972; Porceddu et al., 1973; Pecetti et al., 1992). The lack of allelic similarity between the two centers of diversity can be due to adoption in Ethiopia of a population of landraces from the Middle East that was genetically different from those that can be found there today (founder-migration effect), or as a separate domestication of *T. dicoccum to T. durum*. While both explanations are valid, the tight geographical distribution and low admixture observed among the landraces from the Levantine do not support the hypothesis that a population of landrace existed within this region and then migrated to Ethiopia. However, caution should be used here as one Jordanian landrace was identified among this cluster with 99% of genetic similarity to *T. abyssinicum* types. This landrace does not show the traditional relaxed spike morphology of the *T. abyssinicum* type (**Supplementary Figure S3**) so it remains extremely hard to conclude if this indeed represents the one original landrace population that migrated from the Levantine to Ethiopia. Certainly, its complete lack of genetic similarity to the other Levantine landraces seems to suggest that it was more likely migrated from Ethiopia back to the Levantine, rather than the opposite direction. The second hypothesis then appears slightly more plausible. Domesticated emmer reached Ethiopia some 5,000 years ago (National Research Council, 1996) probably arriving from Egypt along the Silk Road (Luo et al., 2007) and it occupies approximately 7% of the wheat production today under the name of *aja*. Thus, it can be suggested

that Ethiopia is indeed not a secondary center of diversity, but rather a "secondary center of origin," where emmer was further domesticated to durum wheat as it occurred in the Levantine more than 7,000 years before. More targeted study of Ethiopian emmer genetic similarities to landraces from the Levantine and Ethiopia would be need to reach a final conclusion.

Cluster 3 unifies the landraces from the Mediterranean basin (Italy, Greece, Tunisia, Algeria, and Spain), plus few originating from Ethiopia, Afghanistan, Azerbaijan, Kazakhstan, China, Yugoslavia, Iran, and Russia. Foremost, the landrace collected from Ethiopia that belong to this group do not show the typical morphology of *T. abyssinicum* (data not shown). Therefore, it is possible that even if these were collected in Ethiopia, they might not be of the *T. abyssinicum* type. A simple explanation might be that these are rather historical Italian cultivars derived from landraces (Scarascia Mugnozza, 2005) and brought to Ethiopia during the occupation by Italy from 1936 to 1941. Concerning the other countries, it is possible that these landraces are also of Italian origin and from there spread to the neighboring countries through trading in the last millennium. However, with the exclusion of the similarity between Italian and Ethiopian landraces, in the phylogeny tree all genotypes branch out directly from the origin of the branch (**Figure 4B**), which indicates the existence of strong genetic differences among individuals. In fact, this population holds high level of genetic diversity (PIC = 0.17), low levels of alleles fixation (35%), and high level of admixture. Thus, the hypothesis of one single common origin from Italy appears unlikely. In fact, their shared allelic identity suggests that these might have originated from related seed sources, which have then be exposed to similar natural pressures by the environment and accumulated distinct mutations over time.

Cluster 4 is the largest among landraces, the most genetically diverse overall, and it clusters entries from 18 countries (**Supplementary Table S1**). It also shows severe levels of admixtures to the three clusters described above (**Figure 4**). These landraces are therefore likely the result of migration and hybridization of germplasm belonging to clusters 1, 2, and 3. This cluster can be further divided into four sub-populations based on their admixture levels (**Figure 4**). The sub-groups 1 and 2 are derived primarily from Central Asia (Kazakhstan, Afghanistan, India, Armenia, Turkey, Russia, Georgia, and Italy). They are evolutionarily closer to the "Mediterranean types," but maintain good distinction with little similarity among individuals. As described for Cluster 3, this type of similarity with strong individual diversity is probably best explained as multiple separate sampling events from a common seed source, combined with shared environmental pressures. Thus, a scenario can be devised where merchants or tribes departing the Levantine for Central Asia carried with them seeds from neighboring fields. The third sub-group includes entries collected in the Fertile Crescent (Iraq, Afghanistan, and Pakistan), South Asia (Yemen and India), and the Arabian Peninsula (Oman, Saudi Arabia, and Yemen). This group locates evolutionary along the branch of *T. abyssinicum* types. In particular, the landraces from Yemen and Oman are more closely related to the Ethiopian landraces, and are therefore the probable result of dispersion from this secondary center of origin or domestication of durum wheat. The last sub-group is composed of landraces from Iran and Afghanistan located in the same branch with Middle-East landraces, and thus likely dispersed from here.

A special note is required for the landraces of Russian origin. These were identified in clusters 2, 3, and two sub-groups of cluster 4. This level of genetic diversity is normally unexpected for a country so geographically distant from the two centers of origin of durum wheat. However, this vast region has witnessed large migration since its origin, a well-documented source of genetic variation (Vavilov, 1951, 1992).

History of Breeding Exchange and Cross-Hybridization as Explained by the Genetic Diversity of Modern Germplasm

Eleven landraces from the core subset were grouped within the clusters of cultivars and elite lines. The simplest explanation is that these were not true landraces, but rather old tall cultivars that were wrongly labeled during the collecting missions by the gene banks. Alternatively, landraces have often been considered a key resource for contemporary agriculture and thus have been used in plant breeding programs to enlarge the genetic diversity of modern genetic pools (Bradshaw and Ramsay, 2005; Sharma et al., 2013). Hence, it is possible that these landraces are among those utilized in recent years to improve biotic and abiotic tolerance, or were favored by the early breeders like Nazareno Strampelli to develop pure lines (Scarascia Mugnozza, 2005).

Cluster 5 is small, composed of just 13 elite lines and cultivars from ICARDA breeding program and most of them include the cultivar 'Om Rabi' in their pedigree. Om Rabi is the name of the largest river of Morocco, and this name was attributed to one of the first cross ever produced by the ICARDA durum breeding program in 1981, which combined the widely cultivated Jordanian landrace 'Haurani' with the successful CIMMYT line 'Jori69.' Cultivars have been released in 12 countries from this cross under various names ('Cham 5,' 'Tomouh,' 'Om Rabi,' 'Oum Rabi,' 'Omrabi,' 'Gahar,' 'Um Qais,' and 'Aydin93') and they remain widely cultivated by smallholder farmers in the most dry areas of central West Asia and North Africa. Considering that 50% of the genome of this cross is derived from a Levantine landrace, it is not surprising that it shares admixture with Cluster 1, and it has very similar allelic fixation as the landraces from the center of origin of durum wheat (**Table 4**).

Cluster 6 brings together modern and old cultivars developed by Italian breeders. Substantially this set of lines is derived from the initial work of Nazareno Strampelli and the following "fathers" of Italian breeding (Scarascia Mugnozza, 2005). The admixture level is high (**Figure 3**) and it captures 46% of the total alleles assessed (**Table 4**), indicating that several breeding programs worldwide utilized the work carried on in Italy as a base of their cultivar development pipeline. However, the level of genetic diversity is low compared to other breeding clusters (PIC = 0.11), which could be the result of the frequent use in hybridization of a reduced number of founders, in combination

with strong selection pressure for the same traits needed for the Italian growing conditions and the rheological requirements of the pasta industry.

Cluster 7 is located at the center of the graph of the two main IPC by DAPC (**Figure 2**) and it groups together material from several countries and breeding programs such as Spain, Morocco, ICARDA and CIMMYT. It originates from the sharing of several germplasm sources among breeders targeting similar Mediterranean growing conditions. Thus, the genetic similarity between this germplasm can be explained as the common origin of allelic sources, together with the imposition of similar selection pressure for specific traits. This cluster has the highest rate of genetic diversity (PIC = 0.14) and portion of captured alleles (58%) than any other cluster of modern germplasm. In addition, the high admixture (**Figure 3**) and central position in the DAPC graph (**Figure 2**) confirm that this cluster is the founding base that guarantees good exchange of alleles among all other breeding programs.

Germplasm from USA, Australia, and Canada were grouped together in cluster 8, together with four lines from Italy, Spain, and France and two landraces from Algeria. This cluster captures the least amount of available allelic variation (43%) or rare alleles (1%) among breeding programs, and one of the lowest PIC (0.12). Considering the geographical distance between the breeding programs grouped here, and the different environmental conditions, it is a good example of the decay in genetic diversity that other authors have suggested is occurring in breeding programs worldwide (Hoisington et al., 1999; Martos et al., 2005). The tight rheological requirements imposed by the pasta industry has pushed durum wheat breeders to maintain their hybridization programs extremely narrow, using often the same set of standard cultivars as donors of quality traits (Karagöz and Zencirci, 2005; Zencirci and Karagoz, 2005; Altintas et al., 2008). This is reflected by the high number of fixed alleles identified in this cluster (57%) for which genetic diversity no longer exist within these breeding programs. Still, it is important to also indicate that for a large portion of the genome (43%) genetic diversity was captured and can be exploited to make further genetic gain.

Cluster 9 groups together the vast majority of the germplasm of ICARDA included in this analysis, with the exclusion of the 'Om Rabi' derivatives assigned to Cluster 5, and some of the genotypes included in Cluster 7. The durum wheat breeding program of ICARDA officially started in 1981 and run for over 20 years under the umbrella of CIMMYT. This program primarily targeted drylands agriculture using crossing schemes involving both modern and primitive germplasm. It released over the years 100 cultivars in 25 countries (Lantican et al., 2015). Within this group are also included some of the Italian cultivars derived from 'Creso,' a radiation mutant of Strampelli's cultivar 'Capelli,' and several of the CIMMYT lines derived from 'Yavaros,' a CIMMYT cultivar that spread widely in North Africa and is today the most grown in Morocco, Algeria, and Tunisia under the name of 'Karim' (syn. 'Bittern'). The genetic similarity between 'Creso,' ICARDA's and CIMMYT's materials can be found in the pedigree of ICARDA's breeding lines, which widely used 'BiCre' as a parent. In fact, 'BiCre' is derived by the simple cross of 'Bittern' and 'Creso.' This cluster captures 51% of the total allelic diversity available in the panel, and 3% of the rare alleles. Due to the large size ($N = 119$) the PIC is low. This shows that even a breeding program that specifically targets genetic diversity as an adaptation strategy via frequent crosses to primitive germplasm can erode large parts of it by exposing the germplasm to severe selection pressures in challenging environments. Still, this cluster is the second most genetically wide among modern germplasm. Also, ICARDA's breeding lines spread over other two clusters, thus meaning that overall the program was able to maintain acceptable levels of diversity.

The breeding program of CIMMYT has been running for over 50 years. It has had the ability to deliver superior cultivars throughout the developing world and still serves today as source of useful alleles for industrialized countries. As for the ICARDA's cluster, the severe selection pressure during breeding caused a shrinkage in the overall genetic diversity, with the vast majority of the CIMMYT's germplasm clustering mostly in one group (Cluster 10) with the lowest PIC (0.10) and 56% of the genome in fixed status. However, this cluster also captures 21% of the available rare alleles, which is by far the best achievement in that sense among breeding programs. Furthermore, CIMMYT breeding lines can also be found in clusters 9 and 7, which suggest an overall high level of allelic diversity remains available for breeding advancements.

Comparison to Other Population Stratification Studies of Durum Wheat

In previous research, a panel of 190 Spanish durum wheat landraces was attributed to nine sub-populations (Ruiz et al., 2012), while a similar set of the germplasm collection used here, comprising 134 modern durum cultivars, was assigned to six sub-populations by Maccaferri et al. (2003). In our research, the number of clusters used for stratification could have been increased, to allocate four additional sub-populations among landraces of Cluster 4. However, the setting of k is highly dependent on the scope of the research and here the preference was given to capturing similarities rather than divergences. A large portion of admixture among landraces remained unfixed with the set value of k, and this could justify the difference in the number of clusters between our work and that of Ruiz et al. (2012). The ability to distinguish following waves of dispersion among landraces was one of our scopes and this was achieved by finding separation between the two main centers of origin/diversity (Middle East and Ethiopia) and other landraces. Similarly, the division into six clusters of modern material appeared in line with the results of previous authors (Maccaferri et al., 2003) and it provided interesting information about the history of alleles exchange among breeding programs. Slight differences were, however, observed from past works, due to the significant increase in this study of the number of elites derived from the ICARDA breeding program, which alone defined two novel well distinct clusters (5 and 9), and also the study of recent Australian and Canadian cultivars, which also created a cluster not described before by Maccaferri et al. (2003).

Genetic Diversity and the Future of Durum Wheat Breeding

The scenario of today's global cultivation of durum wheat can be summarized in the work of few great breeders: North Africa and the Middle East countries still heavily rely on 'Karim' (syn. 'Bittern,' 'Yavaros79') a mega-cultivar bred by Dr. Gregorio Vazquez (CIMMYT), similarly 'Simeto' is probably the most grown cultivar around the World and it was developed in Italy by Dr. Fortunato Calcagno (Pro.Se.Me), and the 'Cham' series that occupy some of the driest areas of the World were bred at ICARDA in Syria by Dr. Miloudi Nachit. The scenario of the industrialized World is only slightly more segmented, with few mega-cultivars also occupying significant land area. As shown by this genetic diversity study, most of the modern germplasm has only 58 to 44% of the genes still segregating, regardless of the breeding strategy or combination of germplasm utilized. Unexpectedly, the two centers of origin of durum wheat do not appear to be the most exploitable source of allelic diversity with most of their loci in fixed state. Rather, the landraces from Central and South Asia revealed the highest accumulation of rare and normal alleles and should therefore be kept in high consideration for increasing diversity of modern breeding programs. Alternatively, the five clusters of ICARDA, CIMMYT, developed countries, 'Om Rabi' derivatives, and Italian breeding showed limited admixture with each other (**Figure 3**) and therefore their inter-hybridization is a possible source of genetic diversity. Still, this will be possible only if the exchange of seeds for breeding purposes is kept free and unobstructed.

AUTHOR CONTRIBUTIONS

HK, AS, FB, and RO conceived and designed the experiments; HK, AS, and MG performed the laboratory procedures; HK, AS, MG, AA-A, AA, AF-M, BB, FB, and RO analyzed the data; HK and FB wrote the paper; AS, MG, AA-A, AA, AF-M, BB, and RO edited and provided critical review of the manuscript. All authors read and approved the final manuscript.

FUNDING

This research was funded by the Swedish Research Council (Vetenskapsrådet), U-forsk2013, project: Deployment of molecular durum breeding to the Senegal basin: capacity building to face global warming.

ACKNOWLEDGMENTS

The authors wish to thank Prof. R. Tuberosa, Prof. C. Pozniak, Prof. J. Able, Dr. M. Maccaferri, Dr. K. Ammar, and Dr. M. Nachit for sharing some of the germplasm used for this study. We would also like to thank the technical staff of ICARDA for handling field and laboratory procedures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2017.01277/full#supplementary-material

FIGURE S1 | Country of origin of the durum wheat core collection lines (landraces in black, modern lines in gray).

FIGURE S2 | Neighbor-joining tree based on SNPs data using Rogers' genetic distance for 370 lines of durum wheat. Genotypes are coded based on the output of DAPC clustering, and clusters are defined using the vertical black line set at 60% similarity.

FIGURE S3 | Spike morphology of the two centers of origin or diversity of landraces. **(A)** Levantine landraces, **(B)** *T. abyssinicum* type.

TABLE S1 | Complete list with pedigree, origin, IG, and sub-population assignment for each software.

REFERENCES

Acosta-Gallegos, J. A., Kelly, J. D., and Gepts, P. (2007). Pre-breeding and genetic diversity in common bean (*Phaseolus vulgaris*). *Crop Sci.* 47, S44–S59.

Altintas, S., Toklu, F., Kafkas, S., Kilian, B., Brandolini, A., and Ozkan, H. (2008). Estimating genetic diversity in durum and bread wheat cultivars from Turkey using AFLP and SAMPL markers. *Plant Breed.* 127, 9–14.

Autrique, E., Nachit, M. M., Monneveux, P., Tanksley, S. D., and Sorrells, M. E. (1996). Genetic diversity in durum wheat based on RFLPs, morphophysiological traits, and coefficient of parentage. *Crop Sci.* 36, 735–742. doi: 10.2135/cropsci1996.0011183X003600030036x

Bari, A., Street, K., Mackay, M., Endresen, D. T. F., De Pauw, E., and Ahmed, A. (2012). Focused identification of germplasm strategy (FIGS) detects wheat stem rust resistance linked to environmental variables. *Genet. Resour. Crop Evol.* 59, 1465–1481. doi: 10.1007/s10722-011-9775-5

Bonjean, A. P., Angus, W. J., and van Ginkel, M. (2016). *The World Wheat Book: A History of Wheat Breeding*, Vol. 3. Paris: Lavoisier.

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314–331.

Bozzini, A. (1988). "Origin, distribution, and production of durum wheat in the world," in *Durum: Chemistry and Technology*, eds G. Dans Fabriani and C. Lintas (St. Paul, MN: American Association of Cereal Chemists), 1–16.

Bradshaw, J. E., and Ramsay, G. (2005). Utilization of the commonwealth potato collection in potato breeding. *Euphytica* 146, 9–19. doi: 10.1007/s10681-005-3881-4

Cabrera, A., Souza, E., Guttieri, M., Sturbaum, A., Hoffstetter, A., and Sneller, C. (2015). Genetic diversity, linkage disequilibrium and genome evolution in soft winter wheat. *Crop Sci.* 54, 2433–2448. doi: 10.2135/cropsci2013.09.0601

Cantrell, R. G. (1987). Breeding and genetics of durum wheat. *Plant Breed. Rev.* 5, 11–40. doi: 10.1002/9781118061022.ch2

Christiansen, M. J., Andersen, S. B., and Rodomiro, O. (2002). Diversity changes in an intensively bred wheat germplasm during the 20th century. *Mol. Breed.* 9, 1–11. doi: 10.1023/A:1019234323372

Cooper, H. D., Spillane, C., and Hodgkin, T. (2001). *Broadening the Genetic Base of Crop Production*. Wallingford: CABI Publishing.

Delêtre, M., McKey, D. B., and Hodkinson, T. R. (2011). Marriage exchanges, seed exchanges, and the dynamics of manioc diversity. *Proc. Natl. Acad. Sci. U.S.A.* 108, 18249–18254. doi: 10.1073/pnas.1106259108

Deu, M., Sagnard, F., Chantereau, J., Catalayud, C., Hérault, D., Mariac, C., et al. (2013). Diversification of primary gene pool and genetic enhancement of grain legumes. *Front. Plant Sci.* 4:309. doi: 10.3389/fpls.2013.00309

Doyle, J., and Doyle, J. (1987). A rapid procedure for DNA purification from small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.

Dyer, G. A., and Taylor, J. E. (2008). A crop population perspective on maize seed system in Mexico. *Proc. Natl. Acad. Sci. U.S.A.* 105, 470–475. doi: 10.1073/pnas.0706321105

Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1, 47–50.

Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data, linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587.

Feldman, M. (2001). "Origin of cultivated wheat," in *The World Wheat Book: A History of Wheat Breeding*, eds A. P. Dans Bonjean and W. J. Angus (Andover: Intercept Limited), 3–58.

Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 357–374. doi: 10.1146/annurev.arplant.54.031902.134907

Frankel, O. H., Brown, A. H., and Burdon, J. J. (1995). *The Conservation of Plant Biodiversity*. Cambridge: Cambridge University Press.

Gale, M. D., and Youssefian, S. (1985). "Dwarfing genes in wheat," in *Progress in Plant Breeding*, ed. G. E. Russell (London: Butterworths), 1–35.

Gioia, T., Nagel, K. A., Beleggia, R., Fragasso, M., Ficco, D. B. M., Pieruschka, R., et al. (2015). Impact of domestication on the phenotypic architecture of durum wheat under contrasting nitrogen fertilization. *J. Exp. Bot.* 66, 5519–5530. doi: 10.1093/jxb/erv289

Harlan, J. R. (1969). Ethiopia: a center of diversity. *Econ. Bot.* 23, 309–314. doi: 10.1007/BF02860676

Hijmans, R. J., Luigi, G., and Mathur, P. (2012). *DIVA-GIS Version 7.5 Manual.*

Hoisington, D., Khairallah, M., Reeves, T., Ribaut, J. M., Skovmand, B., Taba, S., et al. (1999). Plant genetic resources: What can they contribute toward increased crop productivity? *Proc. Natl. Acad. Sci. U.S.A.* 96, 5937–5943. doi: 10.1073/pnas.96.11.5937

International Wheat Council (1991). *World Production Durum Wheat*. London: The International Wheat Council.

Jombart, T., and Collins, C. (2015). *A Tutorial for Discriminant Analysis of Principal Components (DAPC) Using Adegenet 2.0.*0. London: Imperial College London.

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11:94. doi: 10.1186/1471-2156-11-94

Karagöz, A., and Zencirci, N. (2005). Variation in wheat (Triticum spp.) landraces from different altitudes of three regions of Turkey. *Genet. Resour. Crop Evol.* 52, 775–785. doi: 10.1007/s10722-004-3556-3

Laidò, G., Mangini, G., Taranto, F., Gadaleta, A., Blanco, A., Cattivelli, L., et al. (2013). Genetic diversity and population structure of tetraploid wheats (*Triticum turgidum* L.) estimated by SSR, DArT and pedigree data. *PLoS ONE* 8:e67280. doi: 10.1371/journal.pone.0067280

Lantican, M. A., Payne, T. S., Sonder, K., Singh, R., van Ginkel, M., Baum, M., et al. (2015). *Impacts of International Wheat Improvement Research in the World, 1994-2014*. Mexico: CIMMYT.

LMC International (2009). *Evaluation of Measures Relating to the Durum Wheat Sector within the Context of the Common Agricultural Policy*. Oxford: LMC International.

Luo, M. C., Yang, Z. L., You, F. M., Kawahara, T., Waines, J. G., and Dvorak, J. (2007). The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. *Theor. Appl. Genet.* 114, 947–959. doi: 10.1007/s00122-006-0474-0

Maccaferri, M., Sanguineti, M. C., Donini, P., and Tuberosa, R. (2003). Microsatellite analysis reveals a progressive widening of the genetic basis in the elite durum wheat germplasm. *Theor. Appl. Genet.* 107, 783–797. doi: 10.1007/s00122-003-1319-8

Mackay, M., von Bothmer, R., and Skovmand, B. (2005). Conservation and utilization of plant genetic resources – future directions. *Czech J. Genet. Plant Breed.* 41, 335–344.

MacKey, J. M. (2005). "Wheat: its concept, evolution, and taxonomy," in *Durum Wheat Breeding: Current Approaches and Future Strategies*, Vol. 1, eds C. Royo, M. M. Nachit, N. Di Fonzo, J. L. Araus, W. H. Pfeiffer, and G. A. Slafer (Oxford: Food Products Press), 3–61.

Maier, U. (1996). Morphological studies of free-threshing wheat ears from a Neolithic site in southwest Germany, and the history of the naked wheats. *Veg. Hist. Archaeobot.* 5, 39–55. doi: 10.1007/BF00189434

Martos, V., Royo, C., Rharrabti, Y., and Garcia del Morala, L. F. (2005). Using AFLPs to determine phylogenetic relationships and genetic erosion in durum wheat cultivars released in Italy and Spain throughout the 20th century. *Field Crop Res.* 91, 107–116. doi: 10.1016/j.fcr.2004.06.003

Mengistu, D. K., Kidane, Y. G., Catellani, M., Frascaroli, E., Fadda, C., Pe, M. E., et al. (2016). High-density molecular characterization and association mapping in Ethiopian durum wheat landraces reveals high diversity and potential for wheat breeding. *Plant Biotechnol. J.* 14, 1800–1812. doi: 10.1111/pbi.12538

Mengistu, D. K., Kiros, A. Y., and Pè, M. E. (2015). Phenotypic diversity in Ethiopian durum wheat (*Triticum turgidum* var. durum) landraces. *Crop J.* 3, 190–199. doi: 10.1016/j.cj.2015.04.003

Mohammadi, R., Haghparast, R., Sadeghzadeh, B., Ahmadi, H., Solimani, K., and Amri, A. (2014). Adaptation patterns and yield stability of durum wheat landraces to highland cold rainfed areas of Iran. *Crop Sci.* 54, 944–954. doi: 10.2135/cropsci2013.05.0343

National Research Council (1996). *Lost Crops of Africa: Grains*, Vol. I. Washington, DC: The National Academies Press.

Nesbitt, M., and Samuel, D. (1998). Wheat domestication: archaeobotanical evidence. *Science* 279:1431. doi: 10.1126/science.279.5356.1431e

Novoselović, D., Bentley, A. R., Šimek, R., Dvojković, K., Sorrells, M. E., Gosman, N., et al. (2016). Characterizing Croatian wheat germplasm diversity and structure in a European context by DArT markers. *Front. Plant Sci.* 7:184. doi: 10.3389/fpls.2016.00184

Ortiz, R., Trethowan, R., Ortiz Ferrara, G., Iwanaga, M., Dodds, J. H., Crouch, J. H., et al. (2007). High yield potential, shuttle breeding and a new international wheat improvement strategy. *Euphytica* 157, 365–384. doi: 10.1007/s10681-007-9375-9

Pecetti, L., and Annicchiarico, P. (1998). Agronomic value and plant type of Italian durum wheat cultivars from different eras of breeding. *Euphytica* 99, 9–15. doi: 10.1023/A:1018346901579

Pecetti, L., Annicchiarico, P., and Damania, A. B. (1992). Biodiversity in a germplasm collection of durum wheat. *Euphytica* 60, 229–238. doi: 10.1371/journal.pone.0158883

Perrier, X., Flori, A., and Bonnot, F. (2003). "Data analysis methods," in *Genetic Diversity of Cultivated Tropical Plants*, eds P. Hamon, M. Seguin, X. Perrier, and J. C. Glaszmann (Montpellier: Enfield Science Publishers), 43–76.

Porceddu, E., Perrino, P., and Olita, G. (1973). "Preliminary information on an Ethiopian wheat germplasm collection mission," in *Proceedings of the Symposium on Genetic and Breeding of Durum Wheat*, ed. G. T. Scarascia Mugnozza (Bari: University of Bari), 181–200.

Pritchard, J. K., and Rosenberg, N. A. (1999). Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum. Genet.* 65, 220–228. doi: 10.1086/302449

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multi locus genotype data. *Genetics* 155, 945–959.

Pusadee, T., Jamjod, S., Chiang, Y.-C., Rerkasem, B., and Schaal, B. A. (2009). Genetic structure and isolation by distance in a landrace of Thai rice. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13880–13885. doi: 10.1073/pnas.0906720106

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Raman, H., Stodart, B. J., Cavanagh, C., Mackay, M., Morell, M., Milgate, A., et al. (2010). Molecular diversity and genetic structure of modern and traditional landrace cultivars of wheat (*Triticum aestivum* L.). *Crop Pasture Sci.* 61, 222–229. doi: 10.1071/CP09093

Ranieri, R. (2015). Geography of the durum wheat crop. *Pastaria Int.* 6, 24–36.

Rawashdeh, N. K., Haddad, N. I., Al-Ajlouni, M. M., and Turk, M. A. (2007). Phenotypic diversity of durum wheat (*Triticum durum* Desf.) from Jordan. *Genet. Resour. Crop Evol.* 54, 129–138. doi: 10.1007/s10722-005-2636-3

Rohlf, F. J. (1997). *NTSYS-PC. Numerical Taxonomy and Multivariate Analysis System, Version 2.02e*. New York, NY: Exeter Software.

Royo, C., Elias, E., and Manthey, F. (2009). "Durum wheat breeding," in *Cereals–Handbook of Plant Breeding*, ed. M. J. Carena (Berlin: Springer), 199–226.

Royo, C., Maccaferri, M., Alvaro, F., Moragues, M., Sanguineti, M. C., Tuberosa, R., et al. (2010). Understanding the relationships between genetic and phenotypic structures of a collection of elite durum wheat accessions. *Field Crops Res.* 119, 91–105. doi: 10.1016/j.fcr.2010.06.020

Ruiz, M., Giraldo, P., Royo, C., Villegas, D., Aranzana, M. J., and Carrillo, J. M. (2012). Diversity and genetic structure of a collection of Spanish durum wheat landraces. *Crop Sci.* 52, 2262–2275. doi: 10.1371/journal.pone.0166577

Sakamoto, S., and Fukui, K. (1972). Collection and preliminary observation of cultivated cereals and legumes in Ethiopia. *Kyoto Univ. Afr. Stud.* 7, 181–225.

Scarascia Mugnozza, G. T. (2005). *The Contribution of Italian Wheat Geneticists: From Nazareno Strampelli to Francesco D'Amato*. Rome: Accademia Nazionale delle Scienze.

Schiex, T., de Givry, S., Chabrier, P., and Bouchez, M. (2009). *Carthagène - User Documentation*. Available at: https://mulcyber.toulouse.inra.fr/projects/carthagene/

Sharma, S., Upadhyaya, H. D., Varshney, R. K., and Gowda, C. L. L. (2013). Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. *Front. Plant Sci.* 4:309. doi: 10.3389/fpls.2013.00309

Spillane, C., and Gepts, P. (2001). "Evolutionary and genetic perspectives on the dynamics of crop genepools," in *Broadening the Genetic Base of Crop Production*, eds H. D. Cooper, C. Spillane, and T. Hodgkin (Wallingford: CABI Publishing), 25–70.

Tanksley, S. D., and McCouch, S. R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277, 1063–1066. doi: 10.1126/science.277.5329.1063

Tanno, K., and Willcox, G. (2006). How fast was wild wheat domesticated? *Science* 311:1886.

Taylor, R. D., and Koo, W. W. (2015). 2015 outlook of the U.S. and world wheat industries, 2015-2024. *Agribus. Appl. Econ.* 738, 1–23.

Tester, M., and Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science* 327, 818–822. doi: 10.1126/science.1183700

Vavilov, N. I. (1951). The origin, variation, immunity and breeding of cultivated plants. *Science* 115, 433–434. doi: 10.1094/PHYTO-02-14-0031-R

Vavilov, N. I. (1992). *Origin and Geography of Cultivated Plants*. Cambridge: Cambridge University Press.

Waugh, D. (2010). The silk roads in history. *Expedition* 52, 9–22.

Winfield, M. O., Allen, A. M., Burridge, A. J., Barker, G. L., Benbow, H. R., Wilkinson, P. A., et al. (2016). High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol. J.* 14, 1195–1206. doi: 10.1111/pbi.12485

Zencirci, N., and Karagoz, A. (2005). Effect of developmental stages length on yield and some quality traits of Turkish durum wheat (*Triticum turgidum* L. convar. *durum* (Desf.) Mackey) landraces: Influence of developmental stages length on yield and quality of durum wheat. *Genet. Resour. Crop Evol.* 52, 765–774. doi: 10.1007/s10722-004-6146-5

Zohary, D., Hopf, M., and Weiss, E. (2012). *The Domestication of the Plants in the Old World: the Origin and Spread of Cultivated Plants in West Asia, Europe and Nile Valley*. Oxford: Oxford University Press.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Kabbaj, Sall, Al-Abdallat, Geleta, Amri, Filali-Maltouf, Belkadi, Ortiz and Bassi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 06 May 2020
doi: 10.3389/fgene.2020.00316

Combining QTL Analysis and Genomic Predictions for Four Durum Wheat Populations Under Drought Conditions

Meryem Zaïm[1,2†], Hafssa Kabbaj[1,2†], Zakaria Kehel[2], Gregor Gorjanc[3], Abdelkarim Filali-Maltouf[1], Bouchra Belkadi[1], Miloudi M. Nachit[2] and Filippo M. Bassi[2*]

[1] *Laboratory of Microbiology and Molecular Biology, Faculty of Sciences, Mohammed V University, Rabat, Morocco,* [2] *ICARDA, Biodiversity and Integrated Gene Management, Rabat, Morocco,* [3] *The Roslin Institute, The University of Edinburgh, Edinburgh, United Kingdom*

OPEN ACCESS

Edited by:
Alison Bentley,
National Institute of Agricultural Botany (NIAB), United Kingdom

Reviewed by:
Guanglin He,
Sichuan University, China
Changwei Shao,
Yellow Sea Fisheries Research Institute (CAFS), China

***Correspondence:**
Filippo M. Bassi
f.bassi@cgiar.org

[†] *These authors have contributed equally to this work*

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *21 September 2019*
Accepted: *16 March 2020*
Published: *06 May 2020*

Citation:
Zaïm M, Kabbaj H, Kehel Z, Gorjanc G, Filali-Maltouf A, Belkadi B, Nachit MM and Bassi FM (2020) Combining QTL Analysis and Genomic Predictions for Four Durum Wheat Populations Under Drought Conditions. Front. Genet. 11:316. doi: 10.3389/fgene.2020.00316

Durum wheat is an important crop for the human diet and its consumption is gaining popularity. In order to ensure that durum wheat production maintains the pace with the increase in demand, it is necessary to raise productivity by approximately 1.5% per year. To deliver this level of annual genetic gain the incorporation of molecular strategies has been proposed as a key solution. Here, four RILs populations were used to conduct QTL discovery for grain yield (GY) and 1,000 kernel weight (TKW). A total of 576 individuals were sown at three locations in Morocco and one in Lebanon. These individuals were genotyped by sequencing with 3,202 high-confidence polymorphic markers, to derive a consensus genetic map of 2,705.7 cM, which was used to impute any missing data. Six QTLs were found to be associated with GY and independent from flowering time on chromosomes 2B, 4A, 5B, 7A and 7B, explaining a phenotypic variation (PV) ranging from 4.3 to 13.4%. The same populations were used to train genomic prediction models incorporating the relationship matrix, the genotype by environment interaction, and marker by environment interaction, to reveal significant advantages for models incorporating the marker effect. Using training populations (TP) in full sibs relationships with the validation population (VP) was shown to be the only effective strategy, with accuracies reaching 0.35–0.47 for GY. Reducing the number of markers to 10% of the whole set, and the TP size to 20% resulted in non-significant changes in accuracies. The QTLs identified were also incorporated in the models as fixed effects, showing significant accuracy gain for all four populations. Our results confirm that the prediction accuracy depends considerably on the relatedness between TP and VP, but not on the number of markers and size of TP used. Furthermore, feeding the model with information on markers associated with QTLs increased the overall accuracy.

Keywords: genomic selection, consensus map, drought, imputation, QTL analysis, fixed effect, consensus map, genotyping by sequencing (GBS)

INTRODUCTION

Durum wheat (*Triticum durum* Desf., 2n = 4x = 28, AABB) is grown annually on over 17 million hectares worldwide, and it represents one of the bases of the Mediterranean diet. This region is the largest consumer of durum wheat products and the most significant durum import market (Soriano et al., 2017). The Mediterranean basin is subject to frequent droughts and their occurrence is expected to raise in the near future, with a significant negative effect on crop development and production (Xiao et al., 2018). Breeding for durum genotypes that have an improved yield and tolerance to drought remains one of the most strategic methods to protect the harvest of this crop (Habash et al., 2009; Tadesse et al., 2016; Kuzmanoviæ et al., 2018). The use of genomic models to analyze the main drought adaptation traits can be deployed to significantly accelerate the breeding effort. Genetic linkage map and QTL mapping are useful tools for discovering genomic regions associated with traits of interest (Zhang et al., 2018). However, the significance of the identified QTLs is often linked to the specific parents used and it rarely proved useful for deployment in large scale breeding. One method to control for this error is to perform QTL discovery in multiple populations at the same time. The first step to achieve this is the development of genetic consensus maps that allow to bridge the discovery across populations. In fact, the development of consensus maps has already been shown to not only bridge the information between populations, but also to increase marker density, improve genome coverage, provide a validation of the marker ordering, and reduce markers gaps due to the absence of polymorphism between two parents (Marone et al., 2012; Maccaferri et al., 2014). Multiple genetic linkage maps have already been developed for wheat, and consensus genetic maps have been constructed for hexaploid wheat (Somers et al., 2004; Wang et al., 2014) and durum wheat (Maccaferri et al., 2014, 2015). Furthermore, high-throughput DNA sequencing technologies have now enabled the deployment of reliable and affordable marker coverage via genotype-by-sequencing (GBS), a methodology that relies on restriction enzymes to reduce the amount of genome to be sequenced (Poland et al., 2012; Edae et al., 2017). Numerous recent studies have used this marker system to identify quantitative trait loci (QTL) associated with yield, agronomic traits, and physiologic traits in drought and heat-stressed environments (Acuña-Galindo et al., 2015; Sukumaran et al., 2016; Edae et al., 2017; Hussain et al., 2017; Mwadzingeni et al., 2017; Asif et al., 2018; Bhatta et al., 2018; Roselló et al., 2019), in order to pyramid these QTLs via marker-assisted breeding (Edae et al., 2014).

Genomic selection (GS) builds on the concept of QTL analysis, but it explores the whole genome seeking large and small allelic effects (Bassi et al., 2016). Because of its capacity to better handle complex traits with several small effect alleles such as grain yield (GY), GS is now becoming the methodology of choice for incorporation into breeding strategies (Dekkers and Hospital, 2002; Crosbie et al., 2003; Bassi et al., 2016). GS analyzes jointly all markers to explain the total phenotypic variance through the sum of the markers effects (Meuwissen et al., 2001). Once a model is trained, an effect is assigned to each marker-allele, and the 'genomic estimated breeding value' (GEBVs; Meuwissen et al., 2001) can then be calculated for each individual as the sum of its allelic marker effects. The set of individuals used to train the model has both phenotypic and genotypic available and it is defined as the 'training population' (TP). The set of individuals from which the selection is made is defined as the 'breeding population' (BP), and only genotypic data are collected for it. The 'accuracy' of the predicted GEBV is determined by the correlation between GEBV and the true breeding value (TBV) calculated phenotypically for a 'validation population' (VP), which is genotyped and phenotyped, but not used to train the model. The value for accuracy is used to determine the overall success of the GS approach. Therefore, it is important to maintain a high degree of accuracy, and hence to use a TP that best fits the BP. The degree of relatedness between the two populations is often a good predictor of the accuracy that will be achieved. Cross-validation is used to train and develop the prediction models using different sampling techniques in the TP data sets ahead of estimating the GEBVs in the VP. The idea behind this approach is that breeders can derive predictions of the breeding value of an experimental line even before the line has been tested in the field. In turn, this would allow to make decisions on the use of the lines for yield testing or crossing already during the earlier generations (Crossa et al., 2010; Heffner et al., 2011; Bassi et al., 2016).

However, the integration of QTL analysis and GS remains severely understudied. In the present study, four recombinant inbreed lines (RILs) of durum wheat with different level of relatedness were field tested across environments. QTL analysis was performed for GY and TKW and the same populations were then used to assess different GS models for the two traits. The two methods were then combined by fixing the effect of the marker underlying the QTLs into GS models, to reveal a steep increase in the overall accuracy.

MATERIALS AND METHODS

Mapping Populations

Four F_9-derived RILs mapping populations were obtained by random selection of 200 individual durum spikes from each population at the F_4 generation, followed by single seed descent to F_9. At this generation, the individual plants were sampled for DNA extraction, and the seeds of each individual plant bulked. A different number of individuals for each population was then multiplied and used for yield trial to resemble the typical unbalanced dataset used by breeders. The four durum wheat crosses combining ICARDA's elite lines were: Icamor/Gidara2 (IC; 115 RILs) developed by combining the *Hessian fly* resistance of Icamor (F413J.S/3/Arthur71/Lahn//Blk2/Lahn/4/Quarmal) with the high yield potential of Gidara2 (Stojocri/Omrabi3) (see Bassi et al., 2019 for more details); the second population was Jennah Khetifa/Cham1//T.dicoccoides600545/2*Omrabi5 (DRO; 197 RILs) designed for pyramiding the drought tolerance of the Tunisian landrace Jennah Khetifa, wild emmer, and the ICARDA most successful variety Omrabi; the third population was SW Algia//Gidara1/Cham1 (SW; 93 RILs)

aimed at incorporating the *Septoria tritici* resistance of the Tunisian landrace SW Algia with Gidara1; the fourth population was Omrabi3/Omsnima1//Gidara2 (YG; 145 RILs) aimed at combining drought tolerance and yield potential. As indicated, these populations all have sibling relationships with Omrabi, Cham 1, and Gidara used as parental lines. Additional details are reported in **Table 1**.

Field Trials

Field trials were conducted during the 2014–2015 growing season. The experimental design used at all stations was an augmented complete block design with four common repeated checks, and a block size of 24 entries. The trials were conducted at three drought prone stations in Morocco (**Supplementary Figure S1**): Jemaat Shaim (JSH; 32°21′0′′ N and 8°51′0′′ W), Marchouch (MCH; 33°34′3.1′′ N and 6°38′0.1′′ W) and Sidi el Aidi (SAD; 33°9′36′′ N and 7°24′0′′ W); and one irrigated station in Lebanon: Terbol (TER; 33°48′29′′ N and 35°59′22′′ W) (**Table 2** and **Supplementary Figure S1**). All RILs and their parents were planted in plots of 4.2 m^2 at a seeding rate of 280 plants per m^2. The YG population was planted in MCH, JSH, SAD and TER; the DRO population was also planted in all stations except TER; the IC population was sown in two stations MCH and TER; the SW population in just MCH. Agronomic practices were done following standard procedures, with 80 units of nitrogen provided in 2 equal splits, and 40 units of potassium and phosphorous before planting. Weeds were control by tank mixtures of Derby and Pallas. Days to heading (DTH), days to maturity (DTM), plant height (PLH), and spike density per m^2 (SPK) were recorded in MCH and TER. At maturity, 3 m^2 of the plot were combine harvested and the weight was converted to grain yield as Kg ha^{-1}. At all stations except SAD, 1,000 kernels were weighted on a precision balance to derive 1,000-kernels weight (TKW) and express it in grams (g).

DNA Extraction and Genotyping

Leaf samples obtained from F_9 plants were freeze-dried and used for C-TAB DNA extraction. DNA quality was assessed on agarose gel and it was then equilibrated to 100 ng. The DNA was shipped to the Poland lab at Kansas State University for genotyping by sequencing following the protocol of Poland et al. (2012). Briefly, two restriction enzymes (*Pst*I and *Msp*I) were used for genome complexity reduction, followed by 96-multiplex sequencing by bar coding. Low-quality data filtering was carried out according to the following rules: heterozygous calls not superior to 2%, maximum of 30% missing data, and a minor allele frequency superior to 10%.

Consensus Map Procedure

Individual linkage maps for each population were constructed using the statistical software Carthagene v. 1.2.3 (De Givry et al., 2005) and QTL IciMapping V4.1 (Meng et al., 2015). First, all marker sequences were aligned to the available bread wheat genome assembly (Winfield et al., 2016; The International Wheat Genome Sequencing Consortium [IWGSC], 2018) by BLAST with an identity cut-off of 98% (1 SNP variant) and *E*-value of $5e^{-25}$. The *squeeze* function of Carthagene was used to eliminate markers that were wrongly ordered at LOD of 5 based on the genome alignment, followed by *flip* with window size of seven, LOD of 3, and zero iterations to determine the most plausible order of markers within each window. This framework map contained correctly aligned markers along the map and several unassigned markers. In QTL IciMapping, the framework markers were *anchored* while the unassigned markers were not. The *by anchor order* algorithm was used to assign to the different linkage groups the unassigned markers at a set LOD of 5, and then order them based on the position of the framework markers. This operation was then repeated using the newly developed framework map and reducing the LOD to 3. This methodology defined four individual genetic maps for each population.

The construction of the consensus map was performed chromosome by chromosome using the *consensus map from multiple linkage maps sharing common markers* (CMP) function of QTL IciMapping. First, by re-grouping markers at a distance of less than 20 cM to obtain one group for each chromosome, followed by the *by anchor order* option to measure the genetic distances between markers along the consensus map based on their relative positions on each individual map. Markers were then ordered based on their consensus map position in an Excel file. In several cases, a marker polymorphic in one population might be monomorphic in another. To avoid linkage distortions, the monomorphic scores were set to missing. At this point, imputation was done using AlphaIMpute option HMM (Hickey et al., 2012; Antolin et al., 2017) and confirmed with the BIP function of QTL IciMapping (Zhang et al., 2010).

Data Analysis and QTL Mapping

Statistical analysis of the phenotypic data was performed using the R software version 3.4.3 and Genstat program version 18. Best linear unbiased estimates (BLUEs) were estimated across all environments, assuming fixed effects for the genotype from a linear mixed-effects model using R package *lme4* (Bates et al., 2015; R Core Team, 2017).

The discriminant analysis of principal components (DAPC), was performed using the ‘adegenet’ package 1.4-1 (Jombart et al.,

TABLE 1 | Cluster analysis of the genetic diversity among four mapping populations using discriminant analysis of principal components (DAPC) with k = 4, their pedigrees, and maps features.

Pedigree	Individuals	Markers	Total length (cM)	Marker density (cM/Marker)
IC: Icamor/Gidara2	115	646	1720.1	5.3
DRO: Jennah Khetifa/Cham1// T.dicoccoides600545/2*Omrabi5	197	2291	1922.5	1.2
SW: SW Algia//Gidara1/Cham1	93	1212	1795.3	1.8
YG: Omrabi3/Omsnima1//Gidara2	145	521	1683.8	6.1

TABLE 2 | Description of the field testing environments during the 2014–2015 season.

Code	Site	Country	Coordinates	Altitude (m)	Soil type	Climate	Moisture	Annual rainfall (mm)
MCH15	Marchouch	Morocco	33° 34′ 3.1′′ N, 6° 38′ 0.1′′ W	398	Clay vertisol	Mediterranean/warm temperate	Rainfed	449
SAD15	Sidi el Aydi	Morocco	33° 9′ 36′′ N, 7° 24′ 0′′ W	226	Vertisol	Mediterranean/hot and temperate	Rainfed	237
JSH15	Jemhâa Shaim	Morocco	32° 21′ 0′′ N, 8° 51′ 0′′ W	196	Calcic Cambisols	Hot steppe	Rainfed	270
TER15	Terbol	Lebanon	33° 48′ 29′′ N, 35° 59′ 22′′ W	897	Chromic Vertisols	Mediterranean/temperate	Sprinkle	559

2010) in R studio V 3.4.3 (R Core Team, 2017). With DAPC, the hierarchical clustering among populations was determined by applying the R based package "hclust."

QTLs were searched for each individual population in each individual environment via composite interval mapping (CIM) analysis using R/qtl (Broman et al., 2003). The *cim* function was set to five markers covariates and a window size of 10 cM. LOD thresholds were calculated from QTL IciMapping by BIP functionality using 1,000 permutations with a maximum type 1 error probability of 0.05. Only QTLs that appeared at least in two environments and two populations were considered as valid. The distribution of QTLs and the marker density of the consensus and individual population maps were graphically presented on the fourteen chromosomes of durum wheat by a "Circos plot" using R/shiny application (Yu et al., 2018).

Genomic Prediction Modeling

A total of four genomic models were tested as a first step in this study:

(i) a baseline additive model without interactions of genotypic effect (G), environmental (E) effect, and error (ε) (G+E + ε).
(ii) a baseline multi-environment model (G+E + GxE + ε), which assumed interactions between the G and the E.

In both these models, all the effects were assumed to be random with a normal distribution N(0, σ) where σ is the term variance

(iii) the third model was a marker (M) effect model (G+E + GxE + M + ε), where the genotype effect is substituted by an approximation of the genotype's genomic value expressed as a regression on marker covariates.

In this case the model assumes that the genotype's genomic value follows a normal distribution N(0, G σ_g) where σ_g is the genetic variance and G is genomic relationship matrix.

(iv) the last model is the marker × environment model (G+E + GxE + M + MxE + ε) where the marker effect is composed by an effect common to all environment (main effect) plus a random deviation specific to a particular environment (Lopez-Cruz et al., 2015).

Testing of the different models' accuracies was done using DRO, IG and YG populations independently, and setting as cross-validation 80% of the individuals as TP and 20% as VP. The accuracies within and across environments were then calculated as a measure of good fit. The BGLR package (Pérez and de Los Campos, 2014) was used to run all models above from (i) to (vii) by Bayesian ridge regression (BRR) using 10,000 iterations and 5,000 burn in, and 50 replications (de los Campos et al., 2009, 2013). This model induces homogeneous shrinkage of all marker effects toward zero and yields a Gaussian distribution of marker effects. The 50 replications were used to define statistical differences between model accuracies following a one factor ANOVA.

The GxE + MxE model (iv) was selected and used to test additional hypothesis:

(v) the effect of markers number was investigated by comparing predictions using 100, 80, 60, 40, 20, and 10% of the total marker set in combination with reducing the TP population size to 20, 50, and 75% for GY and TKW. The TP individuals were selected randomly in 50 replications, and one factor ANOVA was used to determine significant differences.

(vi) the prediction accuracy of using half sibs vs. full sibs as TP was compared. Each population was set as TP for all others and itself using the whole population as TP and the whole other population as VP.

(vii) to compare the value of MAS and GS, the prediction accuracy was calculated using 50% as TP and 50% as VP for all markers, only markers associated with major effect QTLs, with 44 and 27 markers for GY and TKW, respectively, and by removing these markers linked to QTLs from the set. The TP individuals were selected randomly in 50 replications, and one factor ANOVA was used to determine significant differences.

(viii) the rr-BLUP package v4.6 (Endelman, 2011) was used to run a mixed model estimating the accuracy gain when using markers underlying the QTLs as fixed effects, and the remaining markers as random effects. For this analysis ten random subsets of 50% TP and 50% VP were selected in each population separately (DRO, IG, SW, and YG). QTL analysis was conducted again for each TP subset following the method described above. Those markers that resulted as underlying QTLs in each TP subset were fixed in the model. One factor ANOVA was run for the ten replicates of each population to determine significant differences.

RESULTS

Phenotypic Evaluation

Analysis of variance (ANOVA) showed significant differences for genetic (G) effect ($p < 0.05$) for all the traits across environments, indicating good levels of phenotypic within each population (**Table 3** and **Figure 1**). The genotype by environment interaction (GxE) effect was also significant ($p < 0.05$). The combined BLUE of TKW and GY differed greatly between the two parental lines of the four populations, displaying a normal distribution within RILs populations (**Figure 1**). Gidara 2 and Jk/Ch1 parents in populations IC, DRO and YG had smaller values of TKW than the average, whereas the Icamor parent in population IC had the maximum value (44 g). Similarly, for GY, Gidara 2 had a smaller value than the average GY, same for the parents Icamor and Younes. Cham1 parent of population DRO and SW had the highest recorded GY of this experiment. The population YG had the highest average TKW and GY. Among the four RILs populations, 50.2 g was the highest value recorded for TKW found in IC, and 3,304 kg ha^{-1} the highest GY for YG.

Individual and Consensus Linkage Maps

The GBS process resulted in 22,117 marker calls. Among these, 4,909 matched the curation criteria and were tentatively ordered via genetic mapping. The individual genetic maps contained 646 polymorphic markers covering 1,720.1 cM for the IC population, 2,291 markers spanned 1,922.5 cM for DRO, 1,212 markers were mapped along 1,795.2 cM in SW, and 521 markers over 1,683.7 cM for YG (**Table 4** and **Supplementary Table S1**). The final consensus map incorporated 3,202 markers assigned to 14 linkage groups corresponding to 1,883 unique loci, and spanned a total genetic distance of 2,705.7 cM, with a density of one marker each 0.85 cM (**Table 4**). The A genome, harbored 1,104 markers, covering a linkage distance of 1,133.8 cM, and the B genome 2,098 markers spanning a linkage distance of 1,572 cM. The largest chromosome was 2B, consisting of 540 markers and covering a genetic length of 243.5 cM, while the smallest chromosome in the map was 4A, covering a genetic length of 101.7 cM and consisting of 209 markers. The average size of markers gaps in the consensus map was 22.1 cM. The consensus map across four populations includes 550 RILS lines. Genetic diversity analysis revealed close kinship between IC and DRO, a lower relatedness with SW, and limited kinship to YG (**Table 1**).

QTL Analysis

The identified genetic and phenotypic variations were combined via QTL analysis across the 550 RILs for all measured traits. Significant QTLs were detected for all traits as summarized in **Figure 2** (**Supplementary Tables S2**, **S3**). A total of 31 QTLs

TABLE 3 | Rate of genetic effect across environments of four populations (IC, DRO, SW, and YG) for DTH, DTM, PLH, SPK, TKW, and GY and genotype by environment interactions (GxE) effects.

Pop	GY across env.		DTH			DTM			PLH			SPK	TKW		
	GxE	G	MCH	SAD	TER	MCH	SAD	TER	MCH	SAD	TER	MCH	MCH	TER	JSH
IC	–	0.93*	0.44*	–	0.74*	0.94*	–	0.85	0.93*	–	0.89*	0.95	0.85*	0.99*	–
DRO	0.45*	0.53*	0.90*	1.00*	–	1.00*	0.86*	–	0.99*	0.98*	–	0.97*	0.95*	–	0.97*
SW	–	0.81*	0.94*	–	–	0.74	–	–	0.96*	–	–	0.89	0.97*	–	–
YG	0.63*	0.36*	0.90*	–	0.93*	0.92*	–	0.76*	1.00*	–	0.89*	0.99*	0.95*	0.93*	0.98*

Significant at 0.05 probability level; –, not available data, GxE, genotype by environment interaction effect; G genetic effect; DTH, days to heading; DTM, days to maturity; PLH, plant height; SPK, spike density; TKW, 1000 kernel weight; MCH, Marchouch; SAD, Sidi el Aidi; TER, Terbol; JSH, JemaatShaim.

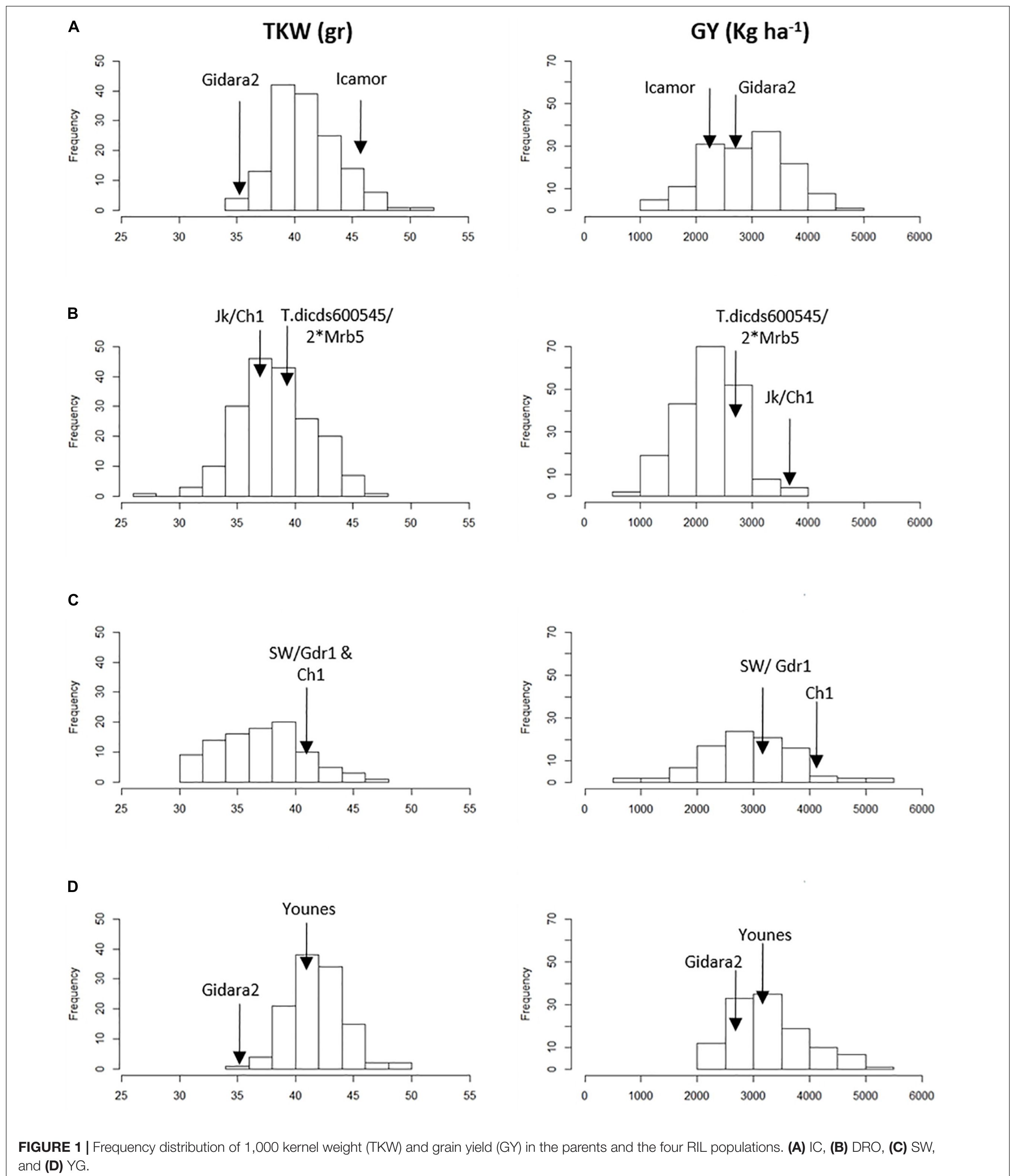

FIGURE 1 | Frequency distribution of 1,000 kernel weight (TKW) and grain yield (GY) in the parents and the four RIL populations. **(A)** IC, **(B)** DRO, **(C)** SW, and **(D)** YG.

were detected across the four populations, explaining from 3.9 to 81.3% of the PV and LOD diverging from 3.7 to 43.5. Six QTLs were found to be associated with GY and independent from the flowering time. In particular, on chromosomes 2B, 4A, and 5B the four independent populations identified consistently the same GY QTL. Six QTLs were detected for TKW on

TABLE 4 | Characteristics of the consensus map.

Chr.	Markers	Loci	Length (cM)	Marker density (cM/Marker)	Size of largest gap (cM)
1A	118	72	138.8	1.2	26.7
1B	257	106	228.5	0.9	24.9
2A	220	164	135.6	0.6	17.9
2B	540	361	243.5	0.5	16.6
3A	146	74	199.5	1.4	21.1
3B	302	189	238.1	0.8	32.6
4A	209	130	101.7	0.5	6.3
4B	197	125	208.3	1.1	29.9
5A	105	38	217.5	2.1	29.7
5B	302	162	245.0	0.8	16.6
6A	162	75	171.8	1.1	17.5
6B	246	155	181.9	0.7	16.6
7A	144	80	168.9	1.2	20.8
7B	254	152	226.6	0.9	31.7
A genome	1104	633	1133.8	1.0	29.7
B genome	2098	1250	1572.0	0.7	32.6

chromosomes 1B, 4B, 6A, 6B, and 7A, explaining 4.7–15.9 of PV and with maximum LOD of 6.1. Interestingly, loci controlling TKW were found to be also associated to GY on chromosome 2B, explaining 8.6 and 4.8% of PV, and LOD of 4.7 and 4.3 respectively.

Genomic Prediction: Identification of the Best Fitting Model (*i, ii, iii, iv*)

Four statistical models (*i, ii, iii, iv*) were tested to determine the best model to be used for each population (**Figure 3**). Non-significant differences could be identified for the IG population with average accuracies that ranged from 0.42 to 0.41. For DRO, the incorporation of the M effect resulted in a significant increase in accuracy from 0.47 to 0.49. The YG population was the most sensitive to the change of model ranging from 0.27 for models without M (*i* and *ii*), to 0.30 for model *iii*, to 0.33 for model *iv* incorporating GxE + MxE. Following these results, the model incorporating GxE + MxE was chosen to be the best suited for all three populations. For the SW population phenotypic data were available only for one environment, therefore a model using only markers effect (*iii*) was used to run genomic predictions for this population.

Genomic Prediction: Effect of Reducing TP and Marker Size (*v*)

The effect of marker number and TP size on prediction accuracies was tested for GY and TKW (v). **Figure 4** shows that when decreasing the number of markers from 3,202 to 320, a slight decrease in prediction accuracies was observed for the different set of TP. For GY, the reduction of markers caused a shift from 0.44 to 0.41 accuracy using 20% of TP, from 0.47 to 0.43 and from 0.49 to 0.44 for 50 and 75% of TP, respectively. For TKW, it dropped from 0.75 to 0.73 and from 0.76 to 0.74 for 20 and 50% of the TP, respectively, while no difference was observed for the 75% of TP between the total number of marker and 10% of it. Statistical analysis revealed no significant differences could be observed when reducing marker number and TP size for any of the two traits.

Genomic Prediction: Importance of Relatedness Between TP and VP (*vi*)

The four populations share common parents and have hence kinship relationships (**Table 1**). It was therefore evaluated if it would be possible to use one population as TP for the others (VP) which have half-sibs relationships. Using TP that were full sibs to the VP resulted in good accuracy values that ranged from 0.35 to 0.47, and from 0.92 to 0.30 for GY and TKW, respectively (**Table 5**). When the TP was not derived from the same cross of the BP (half sibs), the accuracies drop to values close to zero or even negative (**Table 5**). The only acceptable case for GY with an accuracy of 0.29 was obtained when SW was used as TP for IG, but this was not true when IG was used as TP for SW (accuracy of 0.08). The same was observed for TKW, with SW as TP ensuring an accuracy of 0.22, while YG as TP dropped to 0.09 accuracy. Interestingly, the two most genetically related populations, IG and DRO (**Table 1**) also resulted in very poor prediction accuracies when used as TP for each other.

Genomic Prediction: Effect of QTL Analysis on Model Accuracy (*vi, viii*)

Since QTL analysis and GS have been rarely combined, the last objective of this study was to determine if a step of QTL analysis could help improve the GS model's accuracy. A total of 44 and 27 markers were associated via QTL analysis to GY and TKW, respectively (**Figure 2**). To test their value alone, these were used as the only marker to perform genomic predictions and resulted in non-significant accuracies for GY for DRO (0.18), and IG (−0.02), while significant accuracies could be identified for YG (0.29), while an increased was observed for SW (0.54). Similarly, for TKW there was a loss significance for DRO (0.20), IG (0.11) and YG (0.09), while it again increased for SW (0.54) (**Figure 5**). The opposite attempt was also conducted by removing from the whole set all the markers associated with QTLs. In this case the GY and TKW accuracies became non-significant for all populations, except for SW for which it matched what was obtained when using the full marker dataset (**Figure 5**). With the exception of SW, for which the use of only markers associated to QTLs had a positive effect on the prediction accuracies, in all other populations the use of all markers combined was significantly superior.

As it can be expected, the sum of the accuracies of using markers associated to large and small effects does not equal to the accuracy of these combined. It then becomes interesting to assess a model that better incorporates these two by fixing the effect of markers associated to QTLs, while including the random effect of the small impact alleles (viii). To test the suitability to do so in a context that better resembles an actual breeding pipeline, QTL discovery was re-run for each random group of entries composing the TP, and only QTL that could be identified by the specific TP where fixed in the model. **Supplementary**

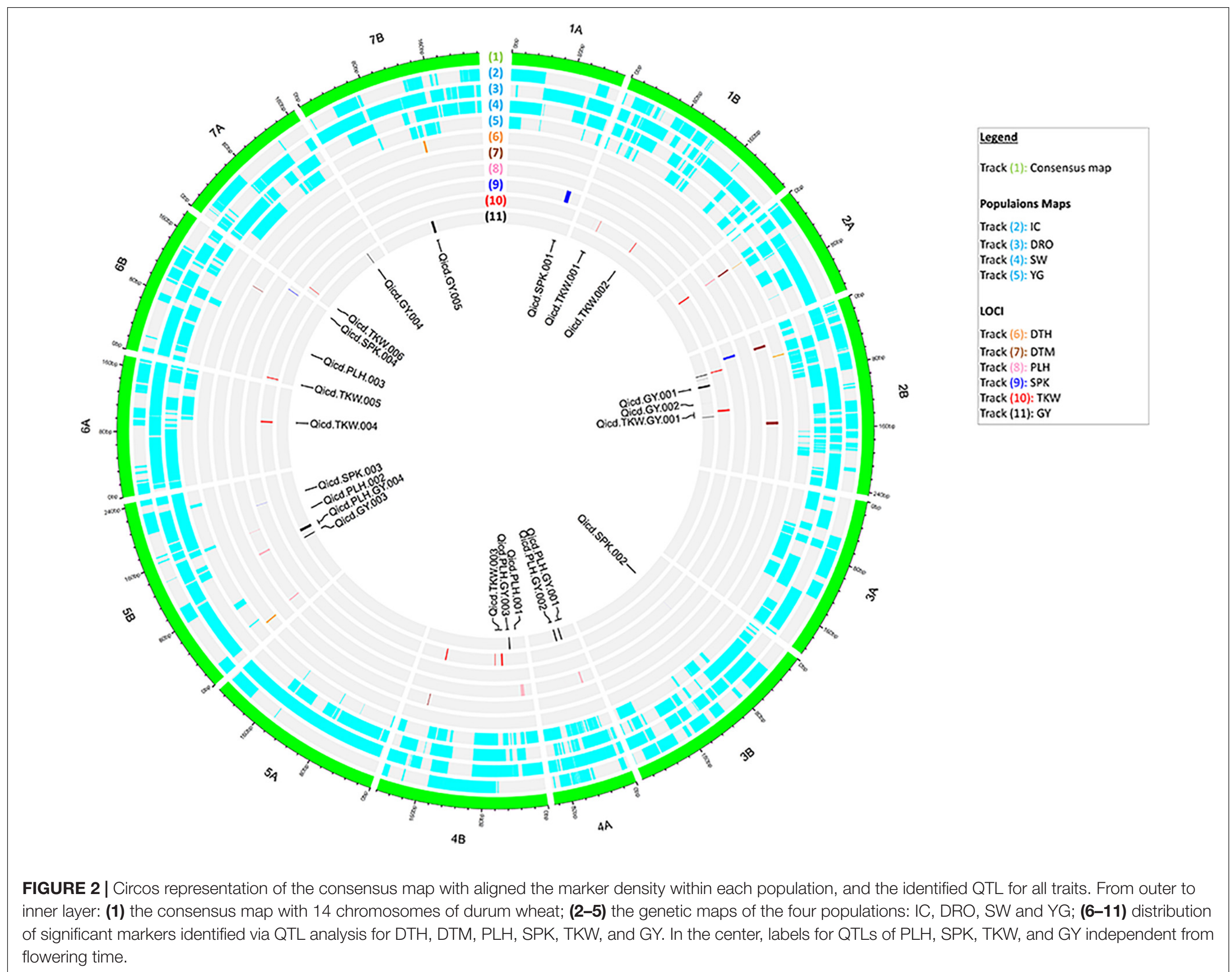

FIGURE 2 | Circos representation of the consensus map with aligned the marker density within each population, and the identified QTL for all traits. From outer to inner layer: **(1)** the consensus map with 14 chromosomes of durum wheat; **(2–5)** the genetic maps of the four populations: IC, DRO, SW and YG; **(6–11)** distribution of significant markers identified via QTL analysis for DTH, DTM, PLH, SPK, TKW, and GY. In the center, labels for QTLs of PLH, SPK, TKW, and GY independent from flowering time.

Table S4 reports how frequently the QTL associated with GY could be re-identified for each TP sub-set. The results of fixing the marker underlying the QTLs in the model is reported in **Figure 6**. For all four populations the accuracies increased significantly ($p < 0.05$) when the QTL-underlying markers were fixed in the model. The average accuracies shifted from 0.35 to 0.47, 0.38 to 0.44, 0.29 to 0.35, and 0.35 to 0.41, for the YG, DRO, IG, and SW populations, respectively. This represents a clear gain of 0.06–0.12 points of accuracy, superior than the 0.01–0.03 obtained by testing different statistical models (*i, ii, iii, iv*).

DISCUSSION

Rapid genetic gain for complex traits via traditional breeding selection is hampered by the difficulty of effectively controlling GxE in the field. Diverting the selection to the use of molecular markers promises to overcome this issue, if adequate models can be defined. Therefore, in our study we deployed four RILs populations that represented well a typical durum wheat breeding program to test the feasibility of replacing phenotypic selection with molecular selection. The four populations showed transgressive segregation when phenotyped for GY and TKW, indicating additive effect loci are present from both parents as it would be expected from a well-designed breeding cross.

A Reliable Consensus Map

To construct a high-density consensus genetic map, a combination of four genetic backgrounds was used by anchoring common markers, followed by imputation of the missing haplotypes. The consensus map of IC, DRO, SW, YG included 14 linkage groups and spanned 2,705 cM, similar to what defined in the four way cross NCCR population map (2,664 cM) of Milner et al. (2016), and the six elite × elite populations durum wheat consensus map (2,631 cM) presented by Maccaferri et al. (2015) and in agreement with other reports ranging from 1,352 cM to 3,598 cM (Blanco et al., 1998; Nachit et al., 2001; Elouafi and Nachit, 2004; Mantovani et al., 2008; Peleg et al., 2008; Patil et al., 2013). The consensus map length was higher by 34% of the average length of the four individual maps. In agreement with

previous studies (Nachit et al., 2001; Elouafi and Nachit, 2004; Peleg et al., 2008; Patil et al., 2013) and contrary to Maccaferri et al. (2015), the A and B genomes had different map lengths, with the B genome (1,572 cM) being longer than A genome (1,133.8 cM). However, similarly to Maccaferri et al. (2015), a smaller number of markers was mapped to the A genome (1,104) compared to the B genome (2,098). The marker density in the consensus map differed along the chromosomes. According to previous studies (Erayman et al., 2004; Saintenac et al., 2011; Maccaferri et al., 2015), this is probably due to the variation of recombination frequency and the potential to accumulate genetic diversity. Markers gaps of 10–33 cM were identified in all chromosomes, except chromosome 4A. Chromosome regions with reduced marker density in 1A, 2A, 3A, and 7A have also been reported in the consensus map of Maccaferri et al. (2014). Overall, the consensus map developed was well in line with previous reported examples and it was hence deemed adequate to perform the targeted study.

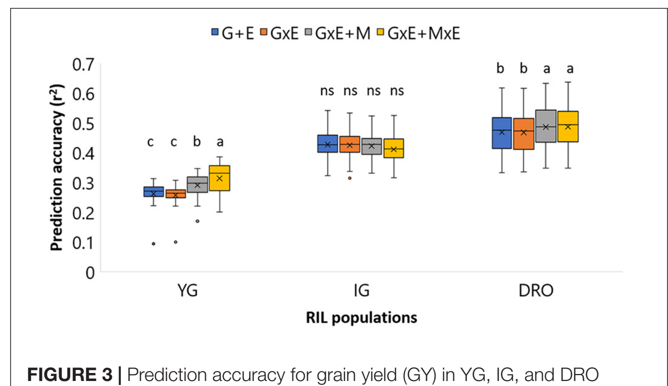

FIGURE 3 | Prediction accuracy for grain yield (GY) in YG, IG, and DRO populations using four different statistical models. G+E, genotype + environment effect; GxE, genotype by environment interaction; GxE + M, genotype by environment interaction + markers effect; GxE + MxE, genotype by environment interaction + markers by environment interaction. The horizontal line represents the average, the square indicates the 2nd and 3rd quartiles, the whiskers represent the 1st and 4th quartiles, the cross the median, and the dots are outliers. The letters indicated classes determined via LSD.

Identification of Major Effect Alleles by QTL Analysis

A total of 31 QTLs were identified for DTH, DTM, PLH, TKW, SPK, and GY, with most of them showing co-localization or pleiotropic effect. Consistent QTLs for GY were detected on chromosomes 2B (Qicd.TKW.DTH.GY.001, Qicd.GY.001, Qicd.GY.002, and Qicd.TKW.GY.001), 4A (Qicd.PLH.GY.001 and Qicd.PLH.GY.002), 4B (Qicd.PLH.GY.003), 5B (Qicd.GY.003 and Qicd.PLH.GY.004), 7A (Qicd.GY.004) and 7B (Qicd.GY.005). Chromosome 2B carries 10 individual QTLs, eight of which were found associated with GY, TKW, and SPK, explaining up to 33.4% of the phenotypic variance. This is in agreement with previous reports on QTLs identified on chromosome 2B associated with GY and its components (Huang et al., 2003; McCartney et al., 2005; Quarrie et al., 2005; Suenaga et al., 2005; Huang et al., 2006; Marza et al., 2006; Maccaferri et al., 2008; Golabadi et al., 2011). Six individual QTLs for TKW were found on chromosomes 1B, 4B, 6A, 6B, and 7A. Except for Qicd.TKW.006 on 7A, which we deem to have been reported here for the first time, the five remaining QTLs have been reported in previous studies by Blanco et al. (2011) and Patil et al. (2013). As indicated by Soriano et al. (2017), QTL influencing SPK were located on chromosomes 2B, 3B, and 5B. Assanga et al. (2017) have also found in winter wheat regions in 1A and 6B that are associated with the same trait.

Major genes associated with phenology were found to have a pleiotropic influence on trait measurement and QTL detection (Acuña-Galindo et al., 2015). Flowering time is a major trait in plant breeding and it provides the basis for plant adaptation. Chromosomes 2A, 2B, 4B, 5B, 6B, and 7B harbored QTLs linked to phenology traits. On 2A and 2B, two clusters of QTLs (Qicd.DTM.PL H.TKW.DTH.001 and Qicd.TKW.DTH.GY.001) were found in approximately the same position corresponding with Ppd-A1 and Ppd-B1 genes defined by several authors (Laurie, 1997; Maccaferri et al., 2008; Wilhelm et al., 2009; Maccaferri et al., 2011; Arjona et al., 2018). In our study, GY was associated to PLH in four QTLs located on chromosomes 4A, 4B, and 5B. Previous studies have also found that PLH genes are strongly associated with QTL for GY and its components (Quarrie et al., 2005; Crossa et al., 2007; Rebetzke et al., 2008; Acuña-Galindo et al., 2015). Borner et al. (2002), Huang et al. (2003, 2006), Blanco et al. (2012), and Patil et al. (2013) found that the short arm of chromosome 2A and its homologous harbor QTL influencing TKW, that was the case for clusters Qicd.DTM.PLH.TKW.DTH.001 and Qicd.TKW.DTH.GY.001. The cluster Qicd.DTM.PLH.TKW.DTH.001 for DTM, DTH, PLH (Soriano et al., 2017) and TKW on chromosome 2A confirms its agronomically important traits contribution as reported in Maccaferri et al. (2011) and Patil et al. (2013). On the homologous region on 2B, the cluster Qicd.TKW.DTH.GY.001 influences DTH, TKW and GY. On chromosome 5B cluster Qicd.DTH.PLH.001 could be related to Vrn-B1 as reported by Hanocq et al. (2004). On the long arm of chromosomes 2B, 4B, 6B, and 7B, the identified QTLs suggest important new regions controlling earliness. Soriano et al. (2017) have also identified a novel QTL on chr. 4B and 7B. In summary, the QTL analysis of these four populations has identified and validated several previously known loci and supports their use for molecular selection.

Selection of the Best Fitting Statistical Models for Genomic Predictions (*i, ii, iii, iv*)

The prediction analysis was conducted on the RILs population using models that account for the relationship matrix (G), environment effect (E), genotype by environment interaction (GxE), markers (M), and marker by environment interaction (MxE). The accuracy of breeding selection using only phenotypic data was computed (**Figure 3**) as G+E and GxE models (*i* and *ii*), to confirm that accuracies of 0.47-0.28 could be obtained via traditional breeding selection for GY. These results confirm what was reported by Crossa et al. (2014): that pedigree (population

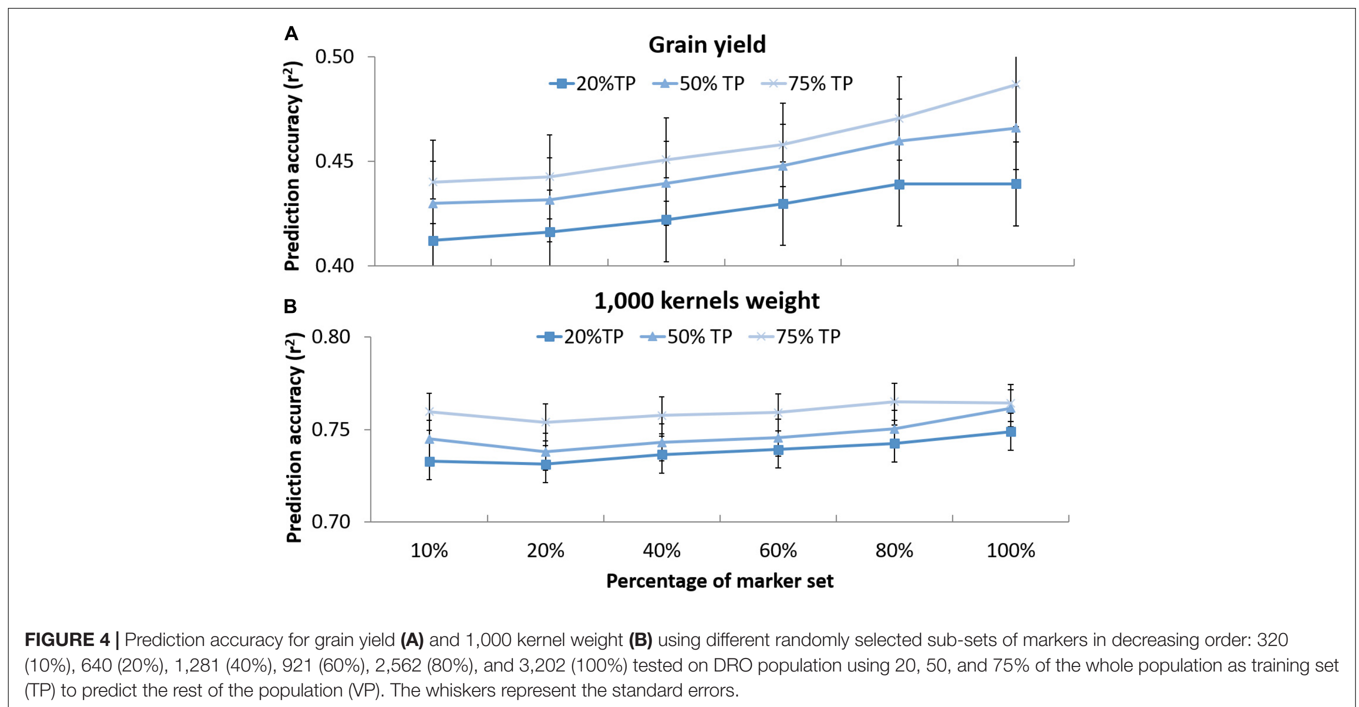

FIGURE 4 | Prediction accuracy for grain yield **(A)** and 1,000 kernel weight **(B)** using different randomly selected sub-sets of markers in decreasing order: 320 (10%), 640 (20%), 1,281 (40%), 921 (60%), 2,562 (80%), and 3,202 (100%) tested on DRO population using 20, 50, and 75% of the whole population as training set (TP) to predict the rest of the population (VP). The whiskers represent the standard errors.

TABLE 5 | Comparison of the prediction accuracies using full sibs and half sibs as training populations for grain yield and 1,000 kernel weight.

	DRO	IG	YG	SW	DRO	IG	YG	SW
	Grain yield				**1,000-kernels weight**			
DRO	0.47	−0.08	−0.11	0.07	0.76	−0.1	0.03	−0.26
IG	−0.09	0.41	0	0.08	−0.08	0.92	−0.02	0.09
YG	−0.07	−0.02	0.35	−0.08	0.12	0	0.83	0.14
SW	0.06	0.29	−0.13	0.37	−0.26	0.22	0.11	0.3

The columns represent the TP and the rows are the BP, the diagonal represents the full sibs relationships.

structure) accounts for a sizeable proportion of the prediction accuracy. These values were set as competitors to determine the success of replacing phenotypic selection with molecular selection. Interestingly, the GS models that incorporated marker effect (*iii, iv*) generated non-significantly different or superior accuracies than traditional breeding selection, indicating a strong role for GS in future breeding (**Figure 3**).

Size and Relatedness of the Training Population (*v, vi*)

Beside academical studies, breeders often have limited resources and tend to reduce costs whenever possible. A decrease in the size of the TP that needs to be both genotyped and phenotyped, and in the number of markers to be used for genotyping can represent important savings (Heffner et al., 2011; Crossa et al., 2014; Bassi et al., 2016). This possibility was tested by varying the proportion of individuals included in TP and VP from 75% TP and 25% VP, which is a very conservative and costly approach, to 50% TP and 50% VP, and even 25% TP and 75% VP. Interestingly, non-significant differences in accuracies could be observed for any of the reductions, for both high and low heritability traits (GY and TKW).

The relatedness between the TP and VP has been identified as a key consideration for predicting complex trait with low heritability. In an ideal scenario, breeders would like to accumulate information for a TP over time, using their normal yield trials as the source for this activity. By logic, the relatedness between such a TP and a BP under selection should be that of half-sibs. To test the feasibility of this approach, the four RIL populations that share half sib relationships were used to predict each other (**Table 5**). This resulted in severe losses of accuracy, reaching values close to zero for both high and low heritable traits (GY and TKW). This is in agreement with Windhausen et al. (2012), who also encountered accuracies close to zero when predicting far-related populations. The relatedness of a TP to the population to be predicted is hence one of the most critical aspect of GS in durum wheat. Therefore, small TP can be effectively deployed to accurately select BP only if these have full sibs relationships with the population to be selected. This is in good agreement with Bassi et al. (2016), who described several breeding schemes to deploy GS in a manner that would allow the TP to

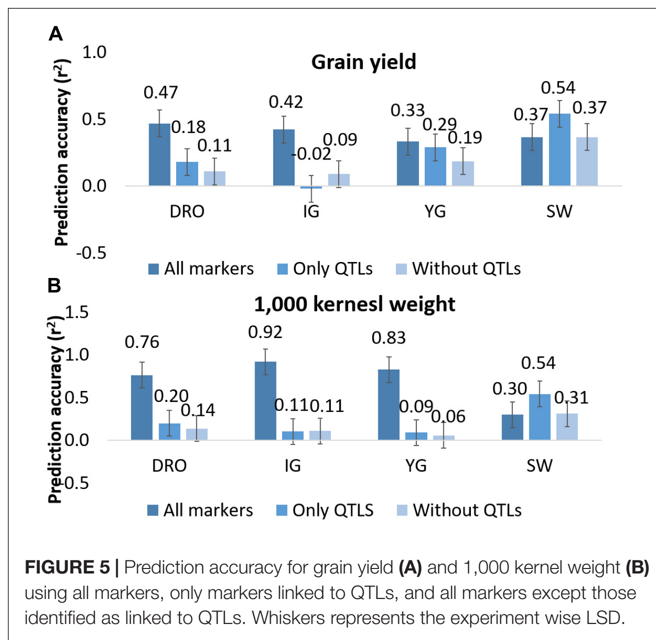

FIGURE 5 | Prediction accuracy for grain yield **(A)** and 1,000 kernel weight **(B)** using all markers, only markers linked to QTLs, and all markers except those identified as linked to QTLs. Whiskers represents the experiment wise LSD.

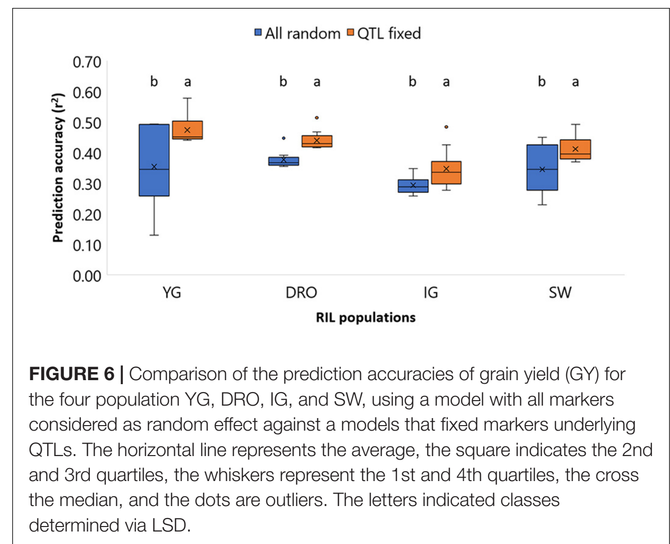

FIGURE 6 | Comparison of the prediction accuracies of grain yield (GY) for the four population YG, DRO, IG, and SW, using a model with all markers considered as random effect against a models that fixed markers underlying QTLs. The horizontal line represents the average, the square indicates the 2nd and 3rd quartiles, the whiskers represent the 1st and 4th quartiles, the cross the median, and the dots are outliers. The letters indicated classes determined via LSD.

be full-sib of the BP under selection, without excessive loss of genetic gain.

Does Markers Number Affect the Predictions? (*v*)

The possibility of deploying GS in breeding is still heavily hindered by the cost associated with genotyping huge populations. A way to reduce the cost of genotyping would be to reduce the number of markers used for the analysis. Here we tested the effect of the markers number to reveal that there was no significant difference in the prediction accuracies between using 3,202 or 320 SNPs as far as the TP and VP are full sibs (**Figure 4**). Hickey et al. (2014) also reported that when using information from related maize bi-parental populations high accuracies can be achieved using a small number of markers. Similarly, Haile et al. (2018) indicated that among advanced durum wheat breeding lines, the reduction from 9,000 to 500 markers did not cause a significant reduction in accuracies. However, it has to be noted that combining a decrease of TP size to 20% of the BP, and 10% of markers number caused the accuracy for GY to drop from 0.48 to 0.41 and for TKW from 0.77 to 0.74. This is a significant reduction of 0.07 and 0.03 points. Still, in the optic of practical application, the values of accuracies remain very close to what achieved using only phenotypic models (G+E and GxE) and hence it could be advisable to deploy small TP and small markers set in breeding if this makes GS a more affordable approach.

Is There an Advantage to Conduct QTL Analysis Before Genomic Predictions? (*vii, viii*)

QTL analysis and GS models rely on the same type of dataset. Therefore, it is of interest to define if there is additive contribution in combining both type of studies. Initially it was tested the effect of using only markers underlying QTLs to make prediction, as a way to simulate a MAS approach (**Figure 5**). The obtained accuracies reached between −0.02 and 0.54, depending on traits and populations. This would suggest that running prediction models using only few markers linked to known genes (44 and 27 for GY and TKW, respectively) could provide some degree of success. For confirmation, the opposite situation was also tested by removing any markers associated to QTL from the whole dataset. Once again, the accuracies dropped significantly for all traits and populations, except for SW. This result suggests that the marker number is not the only factor to ensure high accuracies, but that the ability to define the haplotype of major effect loci is also of critical importance.

The final test was designed to combine the extra information obtained via the definition of major allele effects by QTL analysis with the minor allele effects assessed via GS. Since the initial QTL discovery was conducted using the whole population, while GS models would instead use only sub-set of each population as TP and VP, QTL discovery was re-conducted for each TP subset. All initially identified QTLs were re-identified in 10–50% of the TP subsets (**Supplementary Table S4**) depending on the levels of allelic and phenotypic variation of each random subset. The marker underlying the re-identified QTLs were fixed for each TP subset and used to improve the prediction model. The results are extremely promising, since for all populations the combination of minor allele effects as GS random factor and major allele effects as QTL fixed factor resulted in a significant increase in prediction accuracies. Furthermore, the accuracies value were increased by 0.06–0.12 points, a major increase compared to the 0.02 points of reducing the TP size or changing statistical models. Our results are in partial agreement with Sarinelli et al. (2019) who demonstrated that major genes added as fixed effects always improved model predictive ability, with the greatest gains coming from combinations of multiple genes for days to heading and plant height in a winter wheat panel. Bian and Holland (2017)

also concluded that adding SNPs associated with a given trait as fixed effects resulted in higher predictive abilities when compared to models that only treated SNPs as random effects. Bernardo (2014) pointed out that the prediction accuracy of GS models can be increased by adding major genes as fixed effects when they represent a large proportion of the total variance associated with the trait under consideration ($\geq$10%). Considering that GY remains often the main targeted trait, and also one of the most complex to predict, overall our results support the principle of incorporating fixed effect alleles into a prediction model, especially for markers accounting for a large part of the phenotypic variation. The idea of combining MAS using marker associated to known loci as fixed effects, and all other loci as random effect, becomes interesting for practical breeding applications. Furthermore, there appears to be an additive value in conducting a discovery step via QTL analysis before running genomic predictions, since the additional information can be strategically exploited to increase accuracies.

CONCLUSION

The results of this study provide a framework for better understanding and deploying molecular selection in durum wheat. The use of four populations to define a consensus linkage map allowed the precise identification of significant QTL for agronomic traits. Furthermore, these were incorporated into prediction models to reveal significant gains of accuracy for GY when integrated as fixed effects. Several critical considerations were also tested for their deployment in durum wheat breeding. The results presented here are in good agreement with previous literature and what suggested previously by us for breeding application of GS in wheat (Bassi et al., 2016). In practice, the use of half sibs or distantly related TP does not appear to be an exploitable methodology for GS in durum wheat. Instead, small size full sibs TP needs to be deployed and genotyping costs can be reduced by using just 200–300 SNPs. In addition, known loci linked to traits of interest should be also included in the marker set and used as fixed effects to increase prediction. Most importantly, all genomic prediction models were compared to the accuracy attainable by classical phenotypic selection to confirm that the same results could be achieved via molecular approaches. Altogether, our result provides strong support for the deployment of genomic prediction in durum wheat breeding.

DATA AVAILABILITY STATEMENT

The germplasm described here is available through ICARDA's genebank and can be requested here: https://www.genesys-pgr.org/wiews/SYR002. The genotypic and phenotypic data have been provided as **Supplementary Data Sheet 1**.

AUTHOR CONTRIBUTIONS

MZ, HK, FB, ZK, and GG analyzed the data. AF-M, BB, and MN provided insightful revision and discussions. MZ, HK, FB, and MN produced the data. MZ, HK, and FB wrote the manuscript. All authors reviewed the manuscript.

FUNDING

This study was funded by the Australian Grains Research and Development Corporation (GRDC) project ICA00012: Focused improvement of ICARDA/Australia durum germplasm for abiotic tolerance, while the salary of HK and FB was funded by the Swedish Research Council (Vetenskapsradet) U-Forsk2018 project 2017-05522, "Genomic prediction to deliver heat tolerant wheat to the Senegal River basin: phase II." The genotyping work was covered by the International Treaty on Plant Genetic Resources for Food and Agriculture 2014-2015-2B-PR-02-Jordan: "An Integrated Approach to Identify and Characterize Climate Resilient Wheat for the West Asia and North Africa." Funds for imputation by GG were obtained from BBSRC to The Roslin Institute (BBS/E/D/30002275).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00316/full#supplementary-material

FIGURE S1 | Cartographic location of research stations used for this study. Source: modified from Google Map.

TABLE S1 | Description of the four individual maps issued from IC, DRO, SW and YG populations.

TABLE S2 | Significant QTLs with LOD and phenotypic variance (PV) for the studied traits across environments of the four populations.

TABLE S3 | Significant QTLs with LOD and phenotypic variance (PV) for GY, TKW, SPK and PLH of the four populations.

TABLE S4 | Frequency of re-identifying QTL associated to grain yield in the ten subset of training population.

DATA SHEET S1 | Complete dataset, including phenotyping information for each mapping population presented as BLUEs per location, and full genotyping file for the consensus map.

REFERENCES

Acuña-Galindo, M. A., Mason, R. E., Subramanian, N. K., and Hays, D. B. (2015). Meta-analysis of wheat QTL regions associated with adaptation to drought and heat stress. *Crop Sci.* 55, 477–492. doi: 10.2135/cropsci2013.11.0793

Antolin, R., Nettelblad, C., Gorjanc, G., Money, D., and Hickey, J. M. (2017). A hybrid method for the imputation of genomic data in livestock populations. *Genet. Sel. Evol.* 49:30. doi: 10.1186/s12711-017-0300-y

Arjona, J. M., Royo, C., Dreisigacker, S., Ammar, K., and Villegas, D. (2018). Effect of Ppd-A1 and Ppd-B1 allelic variants on grain number and thousand kernel

weight of durum wheat and their impact on final grain yield. *Front. Plant Sci.* 9:888. doi: 10.3389/fpls.2018.00888

Asif, M. A., Schilling, R. K., and Tilbrook, J., Brien, C., Dowling, K., Rabie, H., et al., (2018). Mapping of novel salt tolerance QTL in an Excalibur × Kukri doubled haploid wheat population. *Theor. Appl. Genet.* 131:2179. doi: 10.1007/s00122-018-3146-y

Assanga, S. O., Fuentealba, M., Zhang, G., Tan, C., Dhakal, S., Rudd, J. C., et al. (2017). Mapping of quantitative trait loci for grain yield and its components in a US popular winter wheat TAM 111 using 90K SNPs. *PLoS One* 12:e0189669. doi: 10.1371/journal.pone.0189669

Bassi, F. M., Bentley, A. R., Charmet, G., and Ortiz, R. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Sci.* 242, 23–36. doi: 10.1016/j.plantsci.2015.08.021

Bassi, F. M., Brahmi, H., Sabraoui, A., Amri, A., Nsarellah, N., Nachit, M. M., et al. (2019). Genetic identification of loci for Hessian fly resistance in durum wheat. *Mol. Breed.* 39:24. doi: 10.1007/s11032-019-0927-1

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bernardo, R. (2014). Genome wide selection when major genes are known. *Crop Sci.* 54, 68–75. doi: 10.2135/cropsci2013.05.0315

Bhatta, M., Morgounov, A., Belamkar, V., and Baenziger, P. S. (2018). Genome-wide association study reveals novel genomic regions for grain yield and yield-related traits in drought-stressed synthetic Hexaploid wheat. *Int. J. Mol Sci.* 19:3011. doi: 10.3390/ijms19103011

Bian, Y., and Holland, J. B. (2017). Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. *Heredity* 118, 585–593. doi: 10.1038/hdy.2017.4

Blanco, A., Bellomo, M. P., Cenci, A., De Giovanni, G., D'Ovidio, R., Lacono, E., et al. (1998). A genetic linkage map of durum wheat. *Theor. Appl. Genet.* 97, 721–728.

Blanco, A., Colasuonno, P., Gadaleta, A., Mangini, G., Schiavulli, A., Simeone, R., et al. (2011). Quantitative trait loci for yellow pigment concentration and individual carotenoid compounds in durum wheat. *Cereal Sci.* 54, 255–264. doi: 10.1016/j.jcs.2011.07.002

Blanco, A., Mangini, G., Giancaspro, A., Giove, S., Colasuonno, P., Simeone, R., et al. (2012). Relationships between grain protein content and grain yield components through quantitative trait locus analyses in a recombinant inbred line population derived from two elite durum wheat cultivars. *Mol. Breeding* 30, 79–92. doi: 10.1007/s11032-011-9600-z

Borner, A., Schumann, E., Furste, A., Coster, H., Leithold, B., Roder, M. S., et al. (2002). Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (*Triticumaestivum* L). *Theor. Appl. Genet.* 105, 921–936. doi: 10.1007/s00122-002-0994-1

Broman, K. W., Wu, H., Sen, S., and Churchill, A. G. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19, 889–890. doi: 10.1093/bioinformatics/btg112

Crosbie, T. M., Eathington, S. R., Johnson, G. R., Edwards, M., Reiter, R., Stark, S., et al. (2003). "Plant breeding: past, present, and future," in *Plant Breeding: The Arnel R. Hallauer International Symposium*, eds K. R. Lamkey and M. Lee (Oxford: Blackwell), 1–50.

Crossa, J., Burgueno, J., Dreisigacker, S., Vargas, M., Herrera-Foessel, S. A., Lillemo, M., et al. (2007). Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics* 177, 1889–1913. doi: 10.1534/genetics.107.078659

Crossa, J., de los Campos, G., Perez, P., Gianola, D., Burgueno, J., Araus, J. L., et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186, 713–724. doi: 10.1534/genetics.110.118521

Crossa, J., Peirez, P., Hickey, J., and BurguenTo, J. (2014). Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* 112, 48–60. doi: 10.1038/hdy.2013.16

De Givry, S., Bouchez, M., Chabrier, P., Milan, D., and Schiex, T. (2005). CARHTAGENE: multipopulation integrated genetic and radiation hybrid mapping. *Bioinformatics* 21, 1703–1704. doi: 10.1093/bioinformatics/bti222

de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., and Calus, M. P. L. (2013). Whole genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193, 327–345. doi: 10.1534/genetics.112.143313

de los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., et al. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182, 375–385. doi: 10.1534/genetics.109.101501

Dekkers, J. C. M., and Hospital, F. (2002). The use of molecular genetics in the improvement of agricultural populations. *Nat. Rev. Genet.* 3, 22–32. doi: 10.1038/nrg701

Edae, E. A., Byrne, P. F., Haley, S. D., Lopes, M. S., and Reynolds, M. P. (2014). Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. *Theor. Appl. Genet.* 127, 791–807. doi: 10.1007/s00122-013-2257-8

Edae, E. A., Olivera, P. D., Jin, Y., and Rouse, M. N. (2017). Genotyping-by-sequencing facilitates a high-density consensus linkage map for *Aegilopsumbellulata*, a wild relative of cultivated wheat. *G3* 7, 1551–1561. doi: 10.1534/g3.117.039966

Elouafi, I., and Nachit, M. M. (2004). A genetic linkage map of the Durum Triticum dicoccoides backcross population based on SSRs and AFLP markers, and QTL analysis for milling traits. *Theor. Appl. Genet.* 108, 401–413. doi: 10.1007/s00122-003-1440-8

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024

Erayman, M., Sandhu, D., Sidhu, D., Dilbirligi, M., and Gill, K. S. (2004). Demarcating the gene-rich regions of the wheat genome. *Nucleic Acids Res.* 32, 3546–3565. doi: 10.1093/nar/gkh639

Golabadi, M., Arzani, A., Maibody, S. M., Tabatabaei, B. S., and Mohammadi, S. (2011). Identification of microsatellite markers linked with yield components under drought stress at terminal growth stages in durum wheat. *Euphytica* 177, 207–221. doi: 10.1007/s10681-010-0242-8

Habash, D. Z., Kehel, Z., and Nachit, M. (2009). Genomic approaches for designing durum wheat ready for climate change with a focus on drought. *Exp. Bot.* 60, 2805–2815. doi: 10.1093/jxb/erp211

Haile, J. K., N'Diaye, A., Clarke, F., Clarke, J., Knox, R., Rutkoski, J., et al. (2018). Genomic selection for grain yield and quality trait in durum wheat. *Mol. Breed.* 38:75. doi: 10.1007/s11032-018-0818-x

Hanocq, E., Niarquin, M., Heumez, E., Rousset, M., and Le Gouis, J. (2004). Detection and mapping of QTL for earliness components in a bread wheat recombinant inbred lines population. *Theor. Appl. Genet.* 110, 106–115. doi: 10.1007/s00122-004-1799-1

Heffner, E., Jannink, J., Iwata, H., Souza, E., and Sorrells, M. (2011). Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Sci.* 51, 2597–2606. doi: 10.2135/cropsci2011.05.0253

Hickey, J., Dreisigacker, S., Crossa, J., Hearne, S., and Babu, R. (2014). Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Sci.* 54, 1476–1488. doi: 10.2135/cropsci2013.03.0195

Hickey, J. M., Kinghorn, B. P., Tier, B., van der Werf, J. H., and Cleveland, M. A. (2012). A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. *Genet. Sel. Evol.* 44:9. doi: 10.1186/1297-9686-44-9

Huang, X. Q., Cloutier, S., Lycar, L., Radovanovic, N., Humphreys, D. G., Noll, J. S., et al. (2006). Molecular detection of QTL for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.). *Theor. Appl. Genet.* 113, 753–766. doi: 10.1007/s00122-006-0346-7

Huang, X. Q., Coster, H., Ganal, M. W., and Roder, M. S. (2003). Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 106, 1379–1389. doi: 10.1007/s00122-002-1179-7

Hussain, W., Baenziger, P. S., Belamkar, V., Guttieri, M. J., Venegas, J. P., Easterly, A., et al. (2017). Genotyping-by-sequencing derived high-density linkage map and its application to QTL mapping of flag leaf traits in bread wheat. *Sci. Rep.* 7:16394. doi: 10.1038/s41598-017-16006-z

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11:94. doi: 10.1186/1471-2156-11-94

Kuzmanoviæ, L., Ruggeri, R., Able, J. A., Bassi, M. F., Maccaferri, M., Tuberosa, R., et al. (2018). Yield performance of chromosomally engineered durum wheat-Thinopyrum ponticum recombinant lines in a range of contrasting rain-fed environments across three countries. *bioRvix [Preprint]* doi: 10.1101/313825

Laurie, D. A. (1997). Comparative genetics of flowering time. *Plant Mol. Biol.* 35, 167–177. doi: 10.1023/A:1005726329248

Lopez-Cruz, M., Crossa, J., Bonnett, D., Dreisigacker, S., Poland, J., Jannink, J. L., et al. (2015). Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. *G3 (Bethesda)* 5, 569–582. doi: 10.1534/g3.114.016097

Maccaferri, M., Cane, M. A., Colalongo, C., Massi, A., Clarke, F., Pozniak, C., et al. (2014). A consensus framework map of durum wheat (*Triticum durum* Desf.) suitable for linkage disequilibrium and genome-wide association mapping. *BMC Genome* 15:873. doi: 10.1186/1471-2164-15-873

Maccaferri, M., Mantovani, P., Tuberosa, R., Deambrogio, E., Giuliani, S., Demontis, A., et al. (2008). A major QTL for durable leaf rust resistance widely exploited in durum wheat breeding programs maps on the distal region of chromosome arm 7BL. *Theor. Appl. Genet.* 117, 1225–1240. doi: 10.1007/s00122-008-0857-5

Maccaferri, M., Ricci, A., Salvi, S., Milner, S. G., Noli, E., Martelli, P. L., et al. (2015). A high-density, SNP-based consensus map of tetraploid wheat as a bridge to integrate durum and bread wheat genomics and breeding. *Plant Biotechnol.* 13, 648–663. doi: 10.1111/pbi.12288

Maccaferri, M., Sanguineti, M. C., Demontis, A., El-Ahmed, A., del Moral, G. L., Maalouf, F., et al. (2011). Association mapping in durum wheat grown across a broad range of water regimes. *Exp. Bot.* 62, 409–438. doi: 10.1093/jxb/erq287

Mantovani, P., Maccaferri, M., Sanguineti, M. C., Tuberosa, R., Catizone, I., Wenzl, P., et al. (2008). An integrated DArT-SSR linkage map of durum wheat. *Mol Breed.* 22, 629–648. doi: 10.1007/s11032-008-9205-3

Marone, D., Laidò, G., Gadaleta, A., Colasuonno, P., Ficco, D. B., Giancaspro, A., et al. (2012). A high-density consensus map of A and B wheat genomes. *Theor. Appl. Genet.* 125, 1619–1638. doi: 10.1007/s00122-012-1939-y

Marza, F., Bai, G. H., Carver, B. F., and Zhou, W. C. (2006). Quantitative trait loci for yield and related traits in the wheat population Ning7840 9 Clark. *Theor. Appl. Genet.* 112, 688–698. doi: 10.1007/s00122-005-0172-3

McCartney, C. A., Somers, D. J., Humphreys, D. G., Lukow, O., Ames, N., Noll, J., et al. (2005). Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452 × 'AC Domain'. *Genome* 48, 870–883. doi: 10.1139/g05-055

Meng, L., Li, H. H., Zhang, L. Y., and Wang, J. K. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop* 3, 269–283. doi: 10.1016/j.cj.2015.01.001

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819.

Milner, S. G., Maccaferri, M., Huang, B. E., Mantovani, P., Massi, A., Frascaroli, E., et al. (2016). A multiparental cross population for mapping QTL for agronomic traits in durum wheat (*Triticum turgidum* ssp. *durum*). *Plant Biotechnol.* 14, 735–748. doi: 10.1111/pbi.12424

Mwadzingeni, L., Shimelis, H., Jasper, D., Rees, G., and Tsilo, T. J. (2017). Genome-wide association analysis of agronomic traits in wheat under drought-stressed and non-stressed conditions. *PLoS One* 12:e0171692. doi: 10.1371/journal.pone.0171692

Nachit, M. M., Elouafi, I., Pagnotta, M. A., El Saleh, A., Lacono, E., Labhilili, M., et al. (2001). Molecular linkage map for an intraspecific recombinant inbred population of durum wheat (*Triticum turgidum* L. var. durum). *Theor. Appl. Genet.* 102, 177–186. doi: 10.1007/s001220051633

Patil, R. M., Tamhankar, S. A., Oak, M. D., Raut, A. L., Honrao, B. K., Rao, V. S., et al. (2013). Mapping of QTL for agronomic traits and kernel characters in durum wheat (*Triticum durum* Desf.). *Euphytica* 190, 117–129. doi: 10.1007/s10681-012-0785-y

Peleg, Z., Saranga, Y., Suprunova, T., Ronin, Y., Roder, M. S., Kilian, A., et al. (2008). High-density genetic map of durum wheat 9 wild emmer wheat based on SSR and DArT markers. *Theor. Appl. Genet.* 117, 103–115. doi: 10.1007/s00122-008-0756-9

Pérez, P., and de Los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198, 483–495. doi: 10.1534/genetics.114.164442

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One* 7:e32253. doi: 10.1371/journal.pone.0032253

Quarrie, S. A., Steed, A., Calestani, C., Semikhodskii, A., Lebreton, C., Chinoy, C., et al. (2005). A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring × SQ1 and its use to compare QTLs for grain yield across a range of environments. *Theor. Appl. Genet.* 110, 865–880. doi: 10.1007/s00122-004-1902-7

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rebetzke, G., Condon, A., Farquhar, G., Appels, R., and Richards, R. (2008). Quantitative trait loci for carbon isotope discrimination are repeatable across environments and wheat mapping populations. *Theor. Appl. Genet.* 118, 123–137. doi: 10.1007/s00122-008-0882-4

Roselló, M., Royo, C., Sanchez-Garcia, M., and Soriano, J. M. (2019). Genetic dissection of the seminal root system architecture in mediterranean durum wheat landraces by genome-wide association study. *Agronomy* 9:364. doi: 10.3390/agronomy9070364

Saintenac, C., Jiang, D., and Akhunov, E. D. (2011). Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biol.* 12:R88. doi: 10.1186/gb-2011-12-9-r88

Sarinelli, J. M., Murphy, J. P., Tyagi, P., Holland, J. B., Johnson, J. W., Mergoum, M., et al. (2019). Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. *Theor. Appl. Genet.* 132, 1247–1261. doi: 10.1007/s00122-019-03276-6

Somers, J. D., Isaac, P., and Edwards, K. (2004). A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 109, 1105–1114. doi: 10.1007/s00122-004-1740-7

Soriano, J. M., Malosetti, M., Rosello, A. M., Sorrells, M. E., and Royo, C. (2017). Dissecting the old *Mediterranean durum* wheat genetic architecture for phenology, biomass and yield formation by association mapping and QTL meta-analysis. *PLoS One* 12:e0178290. doi: 10.1371/journal.pone.0178290

Suenaga, K., Khairallah, M., William, H. M., and Hoisington, D. A. (2005). A new intervarietal linkage map and its application for quantitative trait locus analysis of "gigas" features in bread wheat. *Genome* 48, 65–75. doi: 10.1139/g04-092

Sukumaran, S., Li, X., Zhu, C., Bai, G., Perumal, R., Tuinstra, M. R., et al. (2016). QTL mapping for grain yield, flowering time, and stay-green traits in sorghum with genotyping-by-sequencing markers. *Crop Sci.* 56, 1429–1442. doi: 10.2135/cropsci2015.02.0097

Tadesse, W., Nachit, M., Abdalla, O., and Rajaram, S. (2016). "Wheat breeding at ICARDA: achievements and prospects in the CWANA region," in *The World Wheat Book Volume 3. A History of Wheat Breeding*, eds A. Bonjean, B. Angus, and M. van Ginkel (Paris: Lavoiseier).

The International Wheat Genome Sequencing Consortium [IWGSC] (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191

Wang, S. C., Wong, D. B., Forrest, K., Allen, A., Chao, S. M., Huang, B. E., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol.* 12, 787–796. doi: 10.1111/pbi.12183

Wilhelm, E. P., Turner, A. S., and Laurie, D. A. (2009). Photoperiod insensitive Ppd-A1a mutations in tetraploid wheat (*Triticum durum* Desf.). *Theor Appl Genet.* 118, 285–294. doi: 10.1007/s00122-008-0898-9

Windhausen, V. S., Atlin, G. N., Crossa, J., Hickey, J. M., Grudloyma, P., Terekegne, A., et al. (2012). Effectiveness of genomic prediction of maize

hybrid performance in different breeding populations and environments. *Genes Genomes Genet.* 2, 1427–1436. doi: 10.1534/g3.112.003699

Winfield, M. O., Allen, A. M., Burridge, A. J., Barker, G. L., Benbow, H. R., and Wilkinson, P. A. (2016). High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol.* 14, 1195–1206. doi: 10.1111/pbi.12485

Xiao, D., Bai, H., and Liu, D. L. (2018). Impact of future climate change on wheat production: a simulated case for China's wheat system. *Sustainability* 10:1277. doi: 10.3390/su10041277

Yu, Y., Ouyang, Y., and Yao, W. (2018). ShinyCircos: an R/Shiny application for interactive creation of Circos plot. *Bioinformatics* 34, 1229–1231. doi: 10.1093/bioinformatics/btx763

Zhang, J., Long, Y., Wang, L., Dang, Z., Zhang, T., Song, X., et al. (2018). Consensus genetic linkage map construction and QTL mapping for plant height-related traits in linseed flax (*Linumusitatissimum* L.). *BMC Plant Biol.* 18:160. doi: 10.1186/s12870-018-1366-6

Zhang, L., Liu, D., Guo, X., Yang, W., Sun, J., Wang, D., et al. (2010). Genomic distribution of quantitative trait loci for yield and yield-related traits in common wheat. *Integr. Plant Biol.* 52, 996–1007. doi: 10.1111/j.1744-7909.2010.00967.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zaïm, Kabbaj, Kehel, Gorjanc, Filali-Maltouf, Belkadi, Nachit and Bassi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.