Palacký University of Olomouc

Faculty of Science

Department of Cell Biology and Genetics

and

Institute of Experimental Botany AS CR

Centre of the Region Haná for Biotechnological and Agricultural Research

Elodie Rey

Genomics of Alien Gene Transfer in Wheat

Summary of Ph.D. Thesis

Olomouc

2018

This Ph.D. Thesis was written at the Department of Cell Biology and Genetics, Faculty of Science, Palacký University of Olomouc, between 2013 and 2017.

PhD candidate: Mrs. Elodie Rey

Supervisor: Prof. Ing. Jaroslav Doležel, DrSc.

Reviewers: 1.

2.

3.

The evaluation of this Ph.D. Thesis was written by…….....

……..

The summary of the Ph.D. Thesis was sent for distribution on……

……..

The oral defence will take place on……..before the commission for the Ph.D. study programme Molecular and Cell Biology in the conference room of the Institute of Experimental Botany, Šlechtitelů 31, Olomouc-Holice.

The Ph.D. Thesis is available in biological department library at the Faculty of Science, Palacký University of Olomouc, Šlechtitelů 31, Olomouc-Holice.

Table of Contents

1 Introduction

Bread wheat (*Triticum aestivum* L.) is one of the most important crop species worldwide, providing staple food for 34 % of the human population. It is cultivated on more land than any other crop and its production exceeds 700 million tons per year (http://faostat.fao.org/). As the human population is growing and expected to reach 9.6 billion by 2050, wheat production should increase by about 60 % in order to satisfy the future demand. To meet this challenge, it will be essential to develop wheat cultivars with increased yield, quality and resistance to biotic and abiotic stresses.

Hybridization of bread wheat with its wild relatives is an attractive approach to introgress beneficial genetic variation for agronomical improvement, a process called alien introgression (Feuillet et al., 2008). More than a century of experiments with wide hybridization resulted in numerous combinations between wheat and genomes of its wild relatives, and the introgression of a wide range of beneficial traits into bread wheat (Wulff and Moscou, 2014). However, a few examples of wheat-alien introgression lines have made their way to practical use, mainly due to negative impact of the alien chromatin on the host genome functionality and incomplete genetic compensation affecting agronomical performance (Rey et al., 2015). A better characterization of the interactions between the host genome and the alien chromatin is therefore needed to uncover beneficial and negative effects of alien introgression on wheat phenotype. Identifying different functional levels (spatial, epigenetic, transcriptomic, proteomic) at which host and alien genomes interact will also improve our understanding of the mechanisms underlying alien gene transfer in wheat, and help in identifying ways to make the process more efficient.

In 2017, reference genomes were released for bread wheat (IWGSC, 2018) and barley (Mascher et al., 2017) and important progress has been achieved in sequencing and characterization of genomes of wheat wild relatives (Avni et al., 2017; Luo et al., 2017; Rasheed et al., 2017). These novel resources provide unprecedented templates for detailed genome-wide analysis of the interactions between alien and host genomes in the wheat-alien introgression lines and suggest that wheat-barley crosses may be a suitable model to perform these functional analyses.

This work focuses on the interaction between a host wheat genome and introgressed alien chromatin, at the genome and transcriptome levels. We have chosen a wheat-barley chromosome-arm addition line carrying interesting agronomical traits as a model system (Islam et al., 1981). We have sequenced mRNA of both parents and the addition line to reveal changes in gene expression resulting from intergenomic interactions. Due to the hexaploid nature of the wheat genome and its similarity with the barley genome, one of the difficulties was a reliable discrimination between the expression of homeologous genes of wheat and barley genes in the addition line. An important part of this work was thus to establish a reliable protocol for transcriptomic analysis in the wheat-barley addition line, and the comparison of the expression of wheat and barley genes relative to their expression in their natural background. Following this, the work focused on the analysis of genes showing differential expression in the addition line, within their genomic and evolutionary context in order to infer the effect of structural and functional features on changes in gene expression. This Thesis comprises the first large scale genome-wide transcriptomic analysis of a wheat-barley chromosome-arm addition line and delivers important insights into the transcription of both host and alien genes in an alien introgression line, thereby providing a basis for understanding the interactions between wheat and alien genes in introgression breeding materials.

2 Goals of the Thesis

During the course of this work we made the use of novel sequence resources combined with up-to-date bioinformatics technologies to investigate transcriptomic changes affecting host and alien genes as a result of interactions between the two genomes in a wheat-barley chromosome-arm addition line, which was chosen as a model system because of its potential for wheat improvement. The general aim of this work comprised two main objectives:

i. Establish a protocol for transcriptome analysis in wheat-barley alien introgression lines

Because transcriptomic analysis of a wheat-alien introgression line was never performed at a genome-wide level, the first main objective of the Thesis was to define the best technological platform, plant material, experimental design and transcriptome analysis pipeline for the transcriptomic analysis.

ii. Study the impact of inter-genome interactions on gene expression in wheat-barley 7HL introgression line

Once accurate and sensitive transcriptomic changes were assessed for both wheat and barley genes in the introgression plant material, the second objective of the Thesis was to analyse the changes in gene transcription within their genomic, evolutionary and synteny context in order to better understand by which mechanisms they were affected.

3 Summary of Results

In this thesis, I characterized changes in the expression of host and alien genes in wheat-barley chromosome-arm addition line 7HL. The study was motivated by a desire to understand the interactions between alien chromatin and host genome, which are critical for rational use of alien introgression breeding in wheat improvement programmes, and in other crops in general.

The first goal was to develop a protocol for transcriptome analysis in wheat-barley alien introgression lines. Producing and analysing the transcriptomic data itself represented an important part of this work as it necessitated deep understanding of the bioinformatics and statistics aspects of RNA-seq technology, and conditioned the quality of the results of our study. Accurate assessment of the expression of wheat and barley genes in the context of the hexaploid nature of the wheat genome and its relatedness with barley genome is computationally and statistically demanding due to sequence homology between wheat and barley transcripts. Using the latest genome sequence resources for wheat and barley, a high level of biological replication and more stringent criteria than normally used for RNA-seq analysis, we were able for the first time to assess the expression of about 40% of wheat and barley annotated gene content in a wheat-barley introgression line. The approach allowed the transcript abundance of over 35,000 wheat and nearly 1,000 barley (7HL) genes to be estimated in the wheat-barley 7HL addition line, and the identification of the 960 wheat and 389 barley genes differentially transcribed in the addition line in comparison to their expression in the parent. Our results showed that a large proportion (~42%) of genes on barley 7HL chromosome arm were differentially transcribed in the addition line relative to barley alone, while the expression of only 3% of wheat genes was affected.

The second part of the Thesis was dedicated to the study of changes in wheat host and barley introgressed gene expression within their genomic and syntenic context. The analysis of the distribution of differentially transcribed genes across wheat and barley chromosomes revealed a large deletion (~36Mb) on wheat chromosome arm 7AL in the addition line. This was not expected and the deletion was subsequently identified in all addition lines developed from the wheat cv Chinese Spring / barley cv Betzes combination, and was likely present in the stock of wheat used to generate the disomic addition line set. On the other hand, the barley genes affected were distributed along the entire length

of the 7HL arm, but the ratio of genes experiencing differential transcription proved to be higher at the arm ends: at the proximal end, a majority of the genes affected were up-regulated, while at the distal end, the majority was down-regulated. Finally, although the genes which were differentially transcribed were distributed throughout the wheat genome and along the entire length of chromosome 7HL, implying a level of stochasticism, when subjected to a gene ontology analysis, there was evidence for non-randomness, since only a subset of biological functions was affected. Many of the up-regulated genes on barley 7HL were associated with protein translation, synthesis and modification, while on the other hand, the down-regulated ones were related to reproduction and sexual compatibility. The latter genes may contribute to the reduced fertility of the 7HL addition line in comparison with the other wheat-barley ditelosomic addition lines described by Islam (1983).

The present study identified a set of genes of potential relevance for determining gene networks responsible for interspecific incompatibility between wheat and its related species. This set of genes also provides a resource for investigating the molecular basis of differential transcription, which may involve regulatory sequences, epigenetic changes or spatial organization of chromosome domains in the interphase nucleus. Finally, this pioneering study opens the way for further transcriptomic analysis in other alien introgression lines, including breeding plant material.

4 Conclusions

This Thesis focuses on the interaction between a host wheat genome and introgressed alien chromatin, at the genome and transcriptome levels. In the present work, a method was successfully developed for the study of host and alien genes transcription in the complex polyploid genetic context of wheat-alien introgression lines. The results of this analysis comprise the first large scale genome-wide transcriptomic study of a wheat-barley chromosome-arm addition line, and the investigation of differentially transcribed genes within their genomic, evolutionary and functional context. This work therefore represents a significant advance in the characterization of the transcription of both host and alien genes in an alien introgression line. This pioneering study opens the way for further transcriptomic analysis in other alien introgression lines, including breeding plant material, and provide a basis for understanding the interactions between wheat and alien genes in introgression breeding materials.

5 References

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

Feuillet, C., Langridge, P. and Waugh, R. (2008) Cereal breeding takes a walk on the wild side. *Trends in genetics: TIG* **24**, 24-32.

Islam, A., Shepherd, K.W. and Sparrow, D.H.B. (1981) Isolation and characterization of euplasmic wheat-barley chromosome addition lines. *Heredity* **46**, 161-174.

Islam, A.K.M.R. (1983) Ditelosomic additions of barley chromosomes to wheat. *Proceedings of the sixth International Wheat Genetics Symposium / edited by Sadao Sakamoto*.

IWGSC (2018) Shifting the limits in wheat research and breeding through a fully annotated and anchored reference genome sequence. *Science*.

Luo, M.C., Gu, Y.Q., Puiu, D., Wang, H., Twardziok, S.O., Deal, K.R., Huo, N., Zhu, T., Wang, L., Wang, Y., McGuire, P.E., Liu, S., Long, H., Ramasamy, R.K., Rodriguez, J.C., Van, S.L., Yuan, L., Wang, Z., Xia, Z., Xiao, L., Anderson, O.D., Ouyang, S., Liang, Y., Zimin, A.V., Pertea, G., Qi, P., Bennetzen, J.L., Dai, X., Dawson, M.W., Muller, H.G., Kugler, K., Rivarola-Duarte, L., Spannagl, M., Mayer, K.F.X., Lu, F.H., Bevan, M.W., Leroy, P., Li, P., You, F.M., Sun, Q., Liu, Z., Lyons, E., Wicker, T., Salzberg, S.L., Devos, K.M. and Dvorak, J. (2017) Genome sequence of the progenitor of the wheat D genome Aegilops tauschii. *Nature* **551**, 498-502.

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T., Radchuk, V., Dockter, C., Hedley, P.E., Russell, J., Bayer, M., Ramsay, L., Liu, H., Haberer, G., Zhang, X.Q., Zhang, Q., Barrero, R.A., Li, L., Taudien, S., Groth, M., Felder, M., Hastie, A., Simkova, H., Stankova, H., Vrana, J., Chan, S., Munoz-Amatriain, M., Ounit, R., Wanamaker, S., Bolser, D., Colmsee, C., Schmutzer, T., Aliyeva-Schnorr, L., Grasso, S., Tanskanen, J., Chailyan, A., Sampath, D., Heavens, D., Clissold, L., Cao, S., Chapman, B., Dai, F., Han, Y., Li, H., Li, X., Lin, C., McCooke, J.K., Tan, C., Wang, P., Wang, S., Yin, S., Zhou, G., Poland, J.A., Bellgard, M.I., Borisjuk, L., Houben, A., Dolezel, J., Ayling, S., Lonardi, S., Kersey, P., Langridge, P., Muehlbauer, G.J., Clark, M.D., Caccamo, M., Schulman, A.H., Mayer, K.F.X., Platzer, M., Close, T.J., Scholz, U., Hansson, M., Zhang, G., Braumann, I., Spannagl, M., Li, C., Waugh, R. and Stein, N. (2017) A chromosome conformation capture ordered sequence of the barley genome. *Nature* **544**, 427-433.

Rasheed, A., Mujeeb-Kazi, A., Ogbonnaya, F.C., He, Z. and Rajaram, S. (2017) Wheat genetic resources in the post-genomics era: promise and challenges. *Ann Bot*.

Rey, E., Molnár, I. and Doležel, J. (2015) Genomics of Wild Relatives and Alien Introgressions. In: *Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and Genomics* (Molnár-Láng, M., Ceoloni, C. and Doležel, J. eds), pp. 347-381. Cham: Springer International Publishing.

Wulff, B.B. and Moscou, M.J. (2014) Strategies for transferring resistance into wheat: from wide crosses to GM cassettes. *Frontiers in plant science* **5**, 692.

6 List of author's publications

6.1 Original papers

Rey, E., Molnár, I. and Doležel, J. (2015) Genomics of Wild Relatives and Alien Introgressions. In: *Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and Genomics* (Molnár-Láng, M., Ceoloni, C. and Doležel, J. eds), pp. 347-381. Cham: Springer International Publishing.

Rey, E., Abrouk, M. , Keeble-Gagnère, G. , Karafiátová, M. , Vrána, J. , Balzergue, S. , Soubigou-Taconnat, L. , Brunaud, V. , Martin-Magniette, M. , Endo, T. R., Bartoš, J. , Appels, R. and Doležel, J. (2018), Transcriptome reprogramming due to the introduction of a barley telosome into bread wheat affects more barley genes than wheat. Plant Biotechnol J. Accepted Author Manuscript. doi:10.1111/pbi.12913

6.2 Published abstracts – poster presentations

Rey E., Keeble-Gagnère G., Balzergue S., Soubigou-Taconnat L., Brunaud V., Martin-Magniette ML., Endo T. R., Abrouk M., Bartoš J., Appels R. and Doležel J. Transcriptomic analysis of interactions between host and alien genomes in wheat-barley interspecific introgression. In abstracts of: International Conference on Genomics and Bioinformatics; 2016, Izmir, Turkey.

Rey E., Balzergue S., Soubigou-Taconnat L., Brunaud V., Endo T. R., Abrouk M., Bartoš J. and Doležel J. Genomics of alien gene transfer in wheat with an emphasis on the Transcriptomic analysis performed on barley 7HL genes. In abstracts of: EUCARPIA Cereals Section – European Wheat Aneuploid Conference (EWAC) 2015 International Conference, p.47; 2015, May 24-29. Lublin, Poland.

Rey E., Balzergue S., Soubigou-Taconnat L., Brunaud V., Endo T. R., Bartoš J. and Doležel J. Genomics of alien gene transfer in wheat. In abstracts of: EUCARPIA Cereals Section – ITMI Joint Conference, p.330; 2014, June 29-July 4. Wernigerode, Germany.

6.3 Published abstracts – oral presentations

Rey E., Keeble-Gagnère G., Balzergue S., Soubigou-Taconnat L., Brunaud V., Martin-Magniette ML., Endo T. R., Abrouk M., Bartoš J., Appels R. and Doležel J. The impact of genomic interactions on gene expression in a wheat-barley 7HL addition line. In abstract of: The Plant and Animal Genome (PAG) XXVI Conference. 2017, January 13-17. San Diego, USA.

Rey E., Keeble-Gagnère G., Balzergue S., Soubigou-Taconnat L., Brunaud V., Martin-Magniette ML., Endo T. R., Abrouk M., Bartoš J., Appels R. and Doležel J. A transcriptomic analysis of wheat-barley

7HL-addition line: Genomic interactions between host and alien genomes in products of interspecific hybridization**.** In abstract of: PAG ASIA 2016. 2016, June 6-8. Singapore.

7 Summary (in Czech)

V této práci jsem v linii pšenice nesoucí dlouhé rameno chromozomu 7H ječmene (7HL) charakterizovala změny v expresi genů hostitele a genů ječmene. Cílem práce bylo objasnit interakce mezi vneseným cizím chromatinem a hostitelským genomem, které jsou rozhodující pro efektivní využití introgresních linií ve šlechtění pšenice a jiných zemědělských plodin. Prvním cílem práce bylo vyvinout protokol pro analýzu transkriptomu v introgresní linii pšenice - ječmen. Spolehlivé vyhodnocení exprese genů pšenice a ječmene v kontextu hexaploidního genomu pšenice a jeho příbuznosti s genomem ječmene je kvůli sekvenční homologii transkriptů obou druhů výpočetně a statisticky náročné. Díky přístupu k nejnovějším verzím sekvencí genomu pšenice a ječmene, mnohonásobné biologické replikaci a přísnějším kritériím než ty, které jsou běžně používané pro analýzu dat získaných sekvenováním RNA, bylo v introgresní linii pšenice - ječmen možné vyhodnotit expresi asi 40% anotovaných genů obou druhů. Analýzy transkriptomu ukázaly, že zatímco velká část (~ 42%) genů dlouhého ramene 7HL ječmene má v porovnání s ječmenem v adiční linii změněnou transkripci, pouze asi 3% genů pšenice mají v této linii změněnou expresi. Tyto výsledky otevírají cestu pro analýzu změn genové exprese v jiných introgresních liniích pšenice, včetně šlechtitelských materiálů. Druhá část práce byla zaměřena na studium změn exprese genů hostitelského genomu pšenice a vnesených genů ječmene s ohledem na genomické a syntenické souvislosti. Mimo jiné jsme zjistili, že geny jejichž exprese je v adiční linii změněna, mají rozdílné funkce. Analýzy získaných dat vedly také k odhalení velké delece na dlouhém rameni chromozomu 7A pšenice ve studované introgresní linii. Toto zjištění bylo neočekávané a přítomnost stejné delece byla následně potvrzena ve všech introgresních liních pšenice - ječmen získaných křížením kultivaru pšenice Chinese Spring a kultivaru ječmene Betzes. S velkou pravděpodobností byla tedy tato delece přítomná v linii pšenice použité ke křížení s ječmenem. Naše analýzy identifikovaly soubor genů potenciálně vhodných pro identifikaci genových sítí odpovídajících za mezidruhovou inkompatibilitu pšenice a jejich příbuzných druhů. Tento soubor genů je vhodný pro studium molekulárních mechanismů podílejících se na změnách genové exprese v adičních liniích, které mohou zahrnovat regulační oblasti DNA, epigenetické změny a také změny v prostorovém uspořádání chromatinu v buněčném jádře.

Palacký University of Olomouc

Faculty of Science

Department of Cell Biology and Genetics

and

Institute of Experimental Botany AS CR

Centre of the Region Haná for Biotechnological and Agricultural Research

Elodie Rey

Genomics of Alien Gene Transfer in Wheat

Ph.D. Thesis

Supervisor: Prof. Ing. Jaroslav Doležel, DrSc.

Olomouc

2018

Acknowledgements

I would like to express my most sincere thanks to my supervisor and Head of the laboratory, Prof. Ing. Jaroslav Doležel, DrSc., for giving me the opportunity to perform this work in the Centre of the Plant Structural and Functional Genomics, for his scientific guidance, patience and enthusiasm. I would like to thank all the members of the team for their help and for creating a friendly atmosphere. I would like to especially thank Mgr. Jan Bartoš, Ph.D. for constant support and help during the years of my studies. Finally, I would like to thank my family and friends, and particularly my husband, for their support all along this experience and for believing in me and in this project sometimes more than myself.

Declaration

I hereby declare that I elaborated this Ph.D. Thesis independently under the supervision of Prof. Ing. Jaroslav Doležel, DrSc. using only information sources referred in the chapter References.

.....

This work was supported by the Czech Science Foundation (award P501/12/G090) and by the Ministry of Education, Youth and Sports of the Czech Republic (award LO1204 from the National Program of Sustainability I).

BIBLIOGRAPHIC IDENTIFICATION

Author's name: Elodie Rey

Title: Genomics of Alien Gene Transfer in Wheat

Type of Thesis: Ph.D. Thesis

Department: Cell Biology and Genetics

Supervisor: Prof. Ing. Jaroslav Doležel, DrSc.

The Year of Presentation: 2018

Abstract

Bread wheat (*Triticum aestivum* L.) is one of the most important crop species providing staple food for about one-third of the world's population. Increasing wheat productivity and quality is therefore of a major importance in order to maintain and further improve food security around the world. Hybridization of wheat with its wild relatives is a well-recognized approach for wheat improvement enabling the introgression of favourable alleles, genes and gene complexes, which provide resistance to biotic and abiotic stresses and other beneficial traits. However, post-hybridization interactions between the wheat and alien genomes affect the agronomical performance of the resulting introgression lines and may limit successful application of these materials in agriculture. This work focuses on the interaction between a host wheat genome and introgressed alien chromatin, at the genome and transcriptome levels. We have chosen a wheat-barley chromosome-arm addition line carrying interesting agronomical traits as a model system. We have sequenced mRNA of both parents and the addition line to reveal changes in gene expression resulting from intergenomic interactions. Due to the hexaploid nature of the wheat genome and its similarity with the barley genome, one of the difficulties was a reliable discrimination between the expression of homeologous genes of wheat and barley genes in the addition line. An important part of this work was thus to establish a reliable protocol for transcriptomic analysis in the wheat-barley addition line, and the comparison of the expression of wheat and barley genes relative to their expression in their natural background. Following this,

the work focused on the analysis of genes showing differential expression in the addition line, within their genomic and evolutionary context in order to infer the effect of structural and functional features on changes in gene expression. This Thesis comprises the first large scale genome-wide transcriptomic analysis of a wheat-barley chromosome-arm addition line and delivers important insights into the transcription of both host and alien genes in an alien introgression line, thereby providing a basis for understanding the interactions between wheat and alien genes in introgression breeding materials.

Keywords: bread wheat (*Triticum aestivum* L.), alien introgression, transcriptomics, RNA-sequencing

Number of pages/Appendices: 88/6

Language: English

Table of Contents

1 Introduction

Bread wheat (*Triticum aestivum* L.) is one of the most important crop species worldwide, providing staple food for 34 % of the human population. It is cultivated on more land than any other crop and its production exceeds 700 million tons per year (http://faostat.fao.org/). As the human population is growing and expected to reach 9.6 billion by 2050, wheat production should increase by about 60 % in order to satisfy the future demand. To meet this challenge, it will be essential to develop wheat cultivars with increased yield, quality and resistance to biotic and abiotic stresses.

Hybridization of bread wheat with its wild relatives is an attractive approach to introgress beneficial genetic variation for agronomical improvement, a process called alien introgression. More than a century of experiments with wide hybridization resulted in numerous combinations between wheat and genomes of its wild relatives, and the introgression of a wide range of beneficial traits into bread wheat. However, a few examples of wheat-alien introgression lines have made their way to practical use, mainly due to negative impact of the alien chromatin on the host genome functionality and incomplete genetic compensation affecting agronomical performance. A better characterization of the interactions between the host genome and the alien chromatin is therefore needed to uncover beneficial and negative effects of alien introgression on wheat phenotype. Identifying different functional levels (spatial, epigenetic, transcriptomic, proteomic) at which host and alien genomes interact will also improve our understanding of the mechanisms underlying alien gene transfer in wheat, and help in identifying ways to make the process more efficient.

In this Thesis, I focused on the analysis of the impact of inter-genome interactions on the transcription of host and alien genes in a wheat-barley chromosome arm addition line 7HL. The transcriptomic analysis provided the first insights into the expression of host and alien genes in a wheat-alien introgression line and represents an important step towards the functional characterization of the interactions between host and alien genomes.

2 Literature review

2.1 Wheat: importance and challenges

2.1.1 Global importance of wheat

Wheat, together with barley, is one of the founder crops of agriculture in the Neolithic Near East and Europe, whose wild ancestor's domestication in the region of the Fertile Crescent (Figure 1) marked the transition of human society from hunter-gatherer to a sedentary lifestyle (Kilian et al., 2010). Ultimately, these wild plants became the crops upon which the current civilization developed and expanded worldwide.

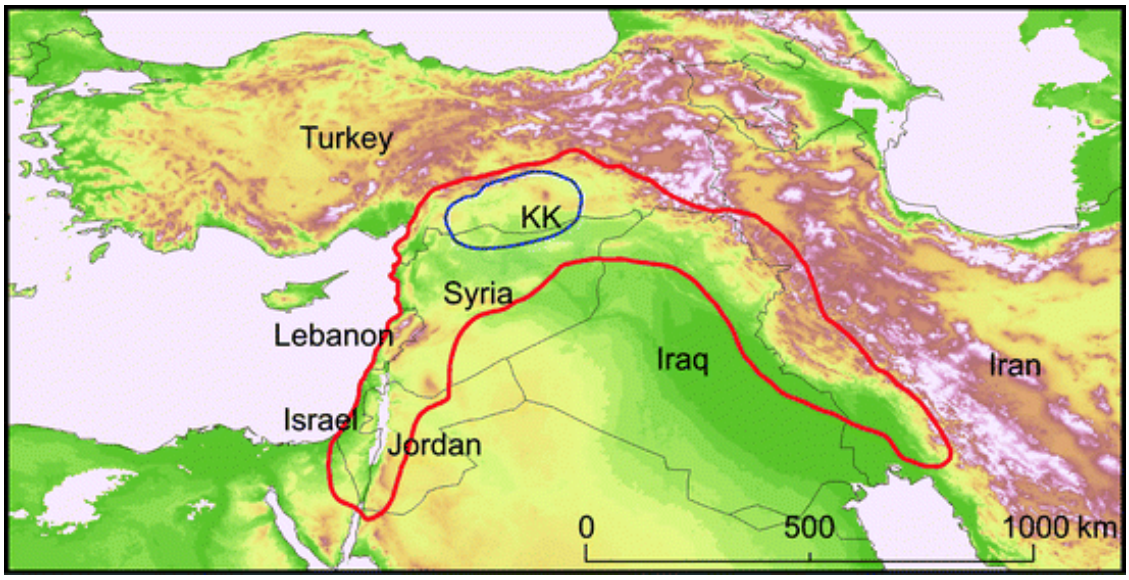


Figure 1: Fertile Crescent (red) and "core area" (blue) of barley (Hordeum vulgare), einkorn wheat (Triticum monococcum) and emmer wheat (Triticum turgidum) domestication within the Fertile Crescent – adapted from (Kilian et al., 2010).

The word 'wheat' defines several wild and cultivated cereal species of the genus *Triticum* (Figure 2), among which the two most importantly grown are pasta wheat (*T. turgidum* ssp. *durum* (Desf.) Husn.) and bread wheat (*T. aestivum* L.). It is cultivated mainly for its grain and its high starch content provides an opportunity for making flour suitable for bread baking, pasta, confectionery, and brewing, among other. Due to its high yield, nutritional value, and adaptability to a wide

range of climatic conditions and environments, wheat has been the basic staple food of the major human civilizations in all continents for more than 8,000 years. Today, it is the most important food grain source for human and is consumed by about 2.5 billion people in 89 countries and provides up to 25% of human calories intake (IWGSC - www.wheatgenome.org).

Figure 2: Cultivated and ancestral wheat species – adapted from (Eversole et al., 2014b)

In 2017, wheat global production reached 743 million tonnes (www.fao.org), and wheat was cultivated on more land area than any other crop (~220 million ha). It is the most widely and extensively grown crop in temperate, Mediterranean-type and sub-tropical parts of both hemispheres, from 67°N in Norway, Finland and Russia, to 45°S in Argentina. Bread wheat is the most widely cultivated wheat representing 95% of grown wheats, with the main producing countries (in terms of both yield and cultivated area) being: EU, China, India, Russia, USA, Canada, Australia, Turkey, Ukraine, Pakistan and Kazakhstan (Figure 3). Thanks to easy transport and storage, it is also an important trading commodity with 174.7 million tonnes of volume and 50 billion dollars of export value (www.faostat.fao.org; data from 2014).

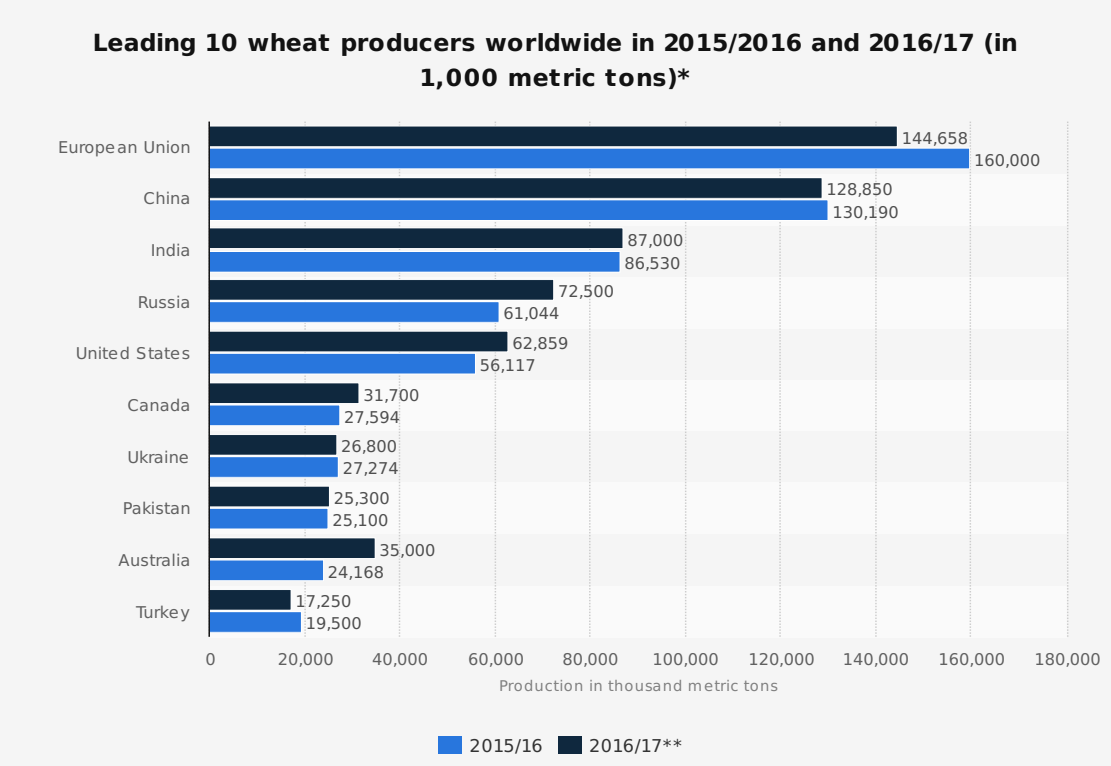


Figure 3: Production of wheat in 2015/2016 and 2016/2017 in million metric tons (MMT) in leading countries (source: www.statista.com and US Department of Agriculture; USDA Foreign Agricultural Service).

2.1.2 Reduced genetic diversity and concerns about food security

Due to global social-economic importance of wheat, it is mandatory to sustain its productivity in order to meet the demand. The latter is expected to increase because of changing diets (amount of food per capita, consumption of meat) and the growth of human population, which is estimated to reach 9.6 billion by 2050 (Tilman et al., 2011; Alexandratos and Bruinsma, 2012). In order to meet the needs of the growing population, cereal production including wheat, maize, barley and rice, must increase by 60% relative to 2005 (Alexandratos and Bruinsma, 2012) (Figure 4).

This increase in production must be achieved in the context of climate change, which brings adverse environmental conditions such as drought, heat and floods, and increased pest pressure. At the same time, the area of arable land is expected to decrease due to competition with habitats and fuel crops, and also soil degradation. Therefore, in order for future crop cultivars to be more productive, they should have higher yields and be more resistant to biotic and abiotic stresses.

Furthermore, as customers now expect to buy healthier products, grown with less inputs, future crop cultivars should also have improved efficiency in fertilizer use (mainly nitrogen and phosphorus) and increased efficiency of photosynthesis.

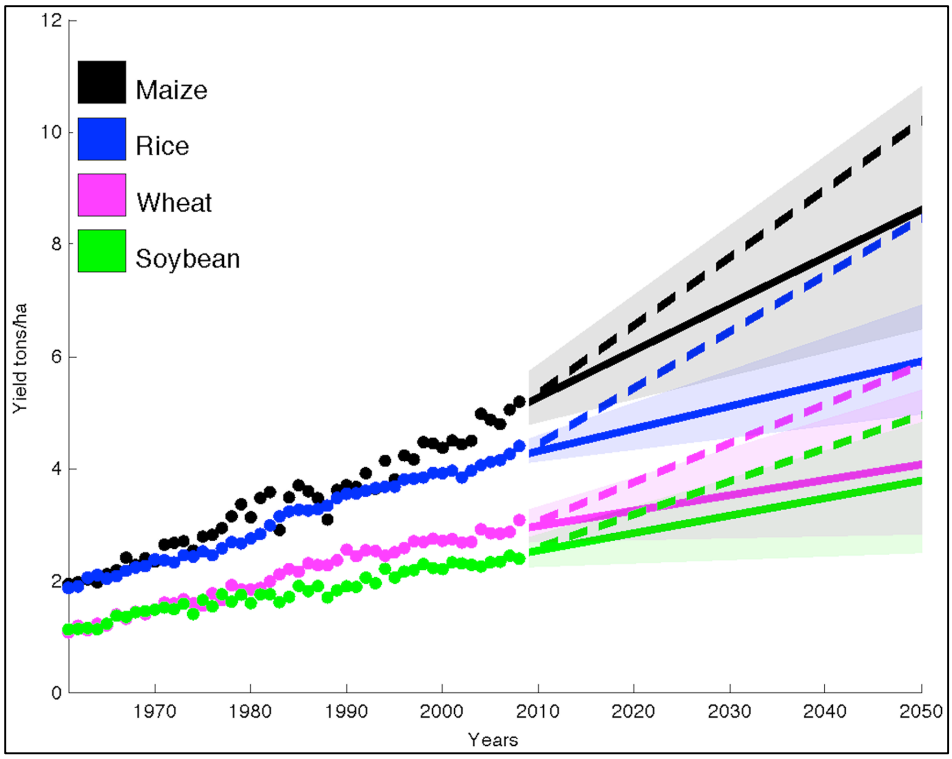


Figure 4: Observed global yield between 1961 and 2008 (closed circles) for maize, rice, wheat, and soybean, and yield projections to 2050 based on the current ~1% genetic gain per year (solid lines) and the ~2.4% yield improvement required each year to double production in these crops by 2050 without bringing additional land under cultivation starting in the base year of 2008 (dash lines) - adapted from (Ray et al., 2013).

Yield increase can be achieved by a combination of new farming practices and the introduction of new cultivars with novel combination of genes and alleles. Currently, the amount of wheat grain yield increases achieved through genetic improvement (also defined as genetic gain) in developed countries averages around 1% per year. In order to meet the demand of the growing population by 2050, 2% genetic gain per year is needed. The low rate of genetic gain results from a limited access to novel alleles in wheat improvement due to narrow genetic diversity of the current elite wheat pool germplasm. Indeed, polyploidization, small number of genotypes that were involved in domestication and are used in current agricultural practices, resulted in genetic bottleneck of cultivated wheat germplasm. The narrow genetic basis reduces the adaptability of

wheat to abiotic stress, increases its susceptibility to biotic pressures, and raises concerns about a possibility to improve wheat performance by the incorporation of novel alleles (Feldman and Levy, 2015). In order to double the genetic gain per year up to 2%, important efforts must be made to recover the genetic diversity in bread wheat pool that was lost during the evolution, domestication and breeding.

2.1.3 Wide hybridization as a key strategy to tackle the wheat genetic bottleneck

For wheat as for other important crops, one of the first options recognized by breeders and geneticists to recover the useful genetic variation that was lost during domestication and intensive breeding, was to introgress important genes and alleles from its wild relatives through wide hybridization. The idea is to combine agronomic performance of bread wheat with beneficial characters of wild relatives, which were not subjected to human selection and thus represent a rich source of diversity for resistance to pathogens and adaptation to various environments.

2.1.3.1 Wide hybridization and alien gene transfer

Alien introgression, also known as introgressive hybridization, is defined as gene transfer from one species into the gene pool of another by repeated backcrossing of an interspecific hybrid with one of its parents. The process of alien gene transfer comprises two main steps:

- **Inter-specific (or inter-generic) cross**, which is the sexual hybridization between wheat and alien species used to bring the 'alien' genome into the wheat host genetic background and which results in generation of amphiploids (hybrids that contain diploid chromosome sets from both parents)
- **Introgression**, which is the incorporation of a part of the alien genome into the host genome, and is achieved by backcrossing the inter-specific hybrid with the host parental genome in order to generate chromosome addition, substitution, translocation, and/or recombinant lines.

The chromosomal insertions can occur either: (i) spontaneously, (ii) through chromosome engineering, including mutagenesis (e.g. ionizing radiation, tissue culture) or the use of

gametocidal chromosome research (Endo, 2007), which aims at inducing fragmentation of donor chromosomes and promoting their recombination with the recipient genome *via* incorrect repair of DNA breaks, or (iii) through manipulation of chromosome pairing during meiosis controlled in wheat by the *Ph1* (*Pairing homeologous 1*) locus (Griffiths et al., 2006) to allow for pairing and recombination between non-homologous chromosomes. Importantly, wide hybridization and alien gene transfer in wheat is facilitated by the polyploid status of wheat, which tolerates chromosomal rearrangements, deletions, insertions and aneuploidy.

2.1.3.2 History of alien gene transfer in wheat

Wheat has a long history of interspecific hybridization, with the first experiments performed by the end of the nineteenth century and beginning of the twentieth century, and which involved hybridization between wheat and rye (Wilson, 1873), wheat and barley (Farrer, 1904), and wheat and *Aegilops* L. (Kihara, 1937). However, larger-scale production of interspecific hybrids was delayed until the introduction of colchicine treatment in 1930s (Blakeslee, 1937), allowing the production of fertile amphiploids by doubling chromosome number in otherwise sterile F1 hybrids. Among other, this provided a way to develop triticale as a new cereal crop. With the advances in hybridization techniques (Kruse, 1973) and establishment of *in vitro* embryo rescue methodology (Murashige and Skoog, 1962), wide hybridization became more successful and the experiments could involve a larger group of wild and cultivated wheat relatives (Mujeeb-Kazi, 1995). Selection and characterization of alien introgressions has been facilitated by methods of molecular cytogenetics, such as *in situ* hybridization (FISH and GISH), and by genotyping using molecular markers (Rayburn and Gill, 1985; Yamamoto and Mukai, 1989; Fedak, 1999; Gupta et al., 1999; Rey et al., 2015). These advances led to the development of a formidable panel of introgression lines of various types and from a wide diversity of wheat relatives (Del Blanco et al., 2001; Warburton et al., 2006; van Ginkel and Ogbonnaya, 2007; Li et al., 2014; Wulff and Moscou, 2014; Mondal et al., 2016) and contributed to the transfer of novel alleles for important traits such as improved yield, resistance to biotic stress, abiotic stress adaptation and

micronutrient availability (Fatih, 1983; Friebe et al., 1996; Reynolds et al., 2001; Wulff and Moscou, 2014; Börner et al., 2015; Mondal et al., 2016).

2.1.3.3 Impact of alien introgression in wheat improvement

Alien introgression has been practiced in wheat improvement programmes for more than a century, and benefits from the large genetic diversity within the tribe Triticeae Dumort. and the extensive wheat genome plasticity. Several important examples in the history of wheat breeding have shown the importance of alien introgression. Perhaps the best example of successful alien introgression into wheat is the rye (*Secale cereale* L.) 1RS chromosome-arm translocation which resulted in the introgression of several genes (*Pm8*/*Sr31*/*Lr26*/*Yr9*) conferring race-specific resistance to pathogens, improved adaptation and stress tolerance, superior aerial biomass and higher kernel weight (Rabinovich, 1998; Zarco-Hernandez et al., 2005; Feuillet et al., 2008; Sharma et al., 2011). Thus, between 1991 and 1995, about 45% of all bread commercially used wheat cultivars carried this translocation (Feuillet et al., 2008) and it is still included in thousands of wheat cultivars across the world. A few other introgressions were generated, mainly with the aim to introgress genes providing resistance to major diseases of wheat, including *Sr36*/*Pm6* from *T. timopheevii* (Zhuk.) Zhuk., *Pm13* from *Aegilops longissima* Schweinf. & Muschl., *Lr28* from *Ae. speltoides* Tausch, *Lr9* from *Ae. umbellulata* Zhuk., *Pch1* and *Sr38*/*Lr37*/*Yr17* from *Ae. ventricosa* Tausch, *Gb2*/*Pm17* from *S. cereale*, and *Lr19*/*Sr25*, *Sr24*/*Lr24*, and *Sr26* from *Lophopyrum elongatum* Löve (Sears, 1956; McIntosh, 1991; Delibes et al., 1993; Friebe et al., 1996). The deployment of these genes has often been on a worldwide scale and contributed to food security over the last century (Wulff and Moscou, 2014). Several review articles underlined the agronomic and economic impact of alien introgressions in leading cultivated wheat varieties around the world and across time, revealing rather a small number of examples, but making a huge impact (Hajjar and Hodgkin, 2007; Gill et al., 2011; Mondal et al., 2016). Nevertheless, the production of useful alien introgressions takes a very long time, is laborious, and often fails due

to various deleterious side effects. As a result, only a small number of wheat-alien introgressions have made their way to agricultural practice.

2.1.3.4 Functional aspects of alien gene transfer

The success of alien introgression is primarily determined by the evolutionary distance between the wild relative and wheat. The obstacles include (i) hybridization incompatibility (pre-embryotic infertility caused by crossability genes, endosperm abnormality and abortion), (ii) negative effects of genes that come along as a linkage block with the alien chromatin carrying the gene of interest (linkage drag), (iii) reduced recombination between the alien and host genomes, (iv) and a less than perfect compensation of alien genes for the loss of wheat genes (Janick, 2008). The limited use of alien introgression lines in wheat breeding results from numerous hurdles faced after the introgression of alien chromatin, which negatively impacts agronomical and quality characteristics of wheat, such as bread making, flour pigmentation and plant fitness (Autrique et al., 1995; Mago et al., 2005). Several studies showed that post-hybridization interactions between the recipient and the donor genomes may affect the function of introgressed and/or host genes and compromise the performance of wheat-alien introgression material. For example, the suppression of a gene determining resistance to a pathogen was shown to be caused by its interaction with the recipient genome (Kerber and Dyck, 1973; Kerber, 1991; Chen et al., 2012). In other cases post-translational modification was shown to impair the function of the donor gene product (Brunner et al., 2011; Stirnweis et al., 2014). According to Hurni et al. (2014), deleterious effects can also result from the formation of a heteromeric complex involving donor and recipient gene products. Due to the lack of characterization of genomes of wheat relatives, negative effects of host and alien genome interactions remain largely unknown and the functional aspects underlying alien gene transfer in wheat remain poorly understood. This paradigm is meant to change with the development of genomic resources for Triticeae species. In 2017, reference genomes were released for bread wheat (IWGSC, 2018) and barley (Mascher et al., 2017) and important progress has been achieved in sequencing and characterization of

genomes of wheat wild relatives (Avni et al., 2017; Luo et al., 2017; Rasheed et al., 2017). These novel resources provide unprecedented templates for detailed genome-wide analysis of the interactions between alien and host genomes in the wheat-alien introgression lines and suggest that wheat-barley crosses may be a suitable model to perform these functional analyses.

2.1.3.5 Importance of wheat-barley introgression lines

Crosses between bread wheat (*T. aestivum*) and cultivated barley (*Hordeum vulgare* L. ssp. *vulgare*), which are the two most important cultivated Triticeae cereals, are attractive in order to incorporate the earliness, favorable amino acid composition, salt and drought tolerance, good tillering ability, stem strength and winter hardiness of barley into wheat (Molnar-Lang et al., 2014).

The first attempts to hybridize the two species date to early 20th century (Farrer 1904), but the first successful cross was reported by the Danish scientist Kruse in 1973 (Kruse, 1973). Barley is one of the distant relatives of wheat within Triticeae tribe, and wheat-barley crosses almost always require *in vitro* techniques for F1 embryo rescue due to lethality and physiological abnormalities or complete sterility. Numerous combinations of wheat-barley genotypes were tested and revealed that the highest seed set was achieved when the wheat cv Chinese Spring was hybridized with barley cv Betzes and cv Ketch (Islam et al., 1976). This is partly due to the crossability genes carried by cv Chinese Spring on the homeologous chromosomes group 5, but should also involve genetic factors from cv Betzes and cv Ketch, currently uncharacterized, whose favourable allelic combinations with wheat genes are responsible for successful hybridization between wheat and barley. The highest seed set is achieved when barley is used as female parent, but this cross direction has little use due to pistilloidy observed in back-cross generation 1 and 2 plants causing the progeny to remain sterile, and preventing development of fertile addition lines (Islam and Shepherd, 1990; Castillo et al., 2014). The reciprocal cross, where wheat is the female and barley the male parent is more difficult to perform (only 1.3 % of successful crosses, compared with 15.4 % in barley-wheat crosses), and only a few laboratories

reported successful crosses (Molnar-Lang et al., 2014).

In 1981, Islam and colleagues developed a series of wheat-barley disomic (Islam et al., 1981) and ditelosomic (Islam, 1983) addition lines from a cross between wheat cv Chinese spring and barley cv Betzes, with wheat as female parent. Each addition line contains a complete chromosome set of wheat and a single pair of barley chromosomes (arms). These wheat-barley disomic and ditelosomic addition lines have been extensively used to physically map genes on barley chromosomes (Cho et al., 2006; Bilgic et al., 2007; Kato et al., 2008; Tang et al., 2011). Efforts were also made to produce wheat-barley crosses using other cultivars, and among the notable successes was a series of chromosome addition lines reported from hybrids between winter wheat line Mv9kr1 and the German two-rowed winter barley cv Igri (Szakacs and Molnar-Lang, 2007), or the six-rowed Ukrainian winter barley cv Manas (Molnar-Lang et al., 2012), a cultivar that is agronomically much better adapted to Central European environmental conditions than the two-rowed spring barley cv Betzes. The wheat-barley addition lines produced in various cultivar combinations (Chinese Spring × Betzes, Mv9 kr1 × Igri, Asakaze × Manas) had several morphological traits in common (Molnar-Lang et al., 2012).

Among the interesting traits that were identified when barley chromosomes (H) were introduced to wheat, Murai et al. (1997) described early flowering time in Chinese Spring-Betzes addition line 5H. In the absence of 5H addition for the Mv9kr1-Igri and Asakaze-Manas wheat/winter barley addition lines, 7H chromosome addition lines were found to be about 12 days earlier than the late flowering 4H addition lines (Farkas et al., 2014), which was later found to be contributed by 7HL telosome (Turkosi et al., 2016). On the other hand, genes shown to be involved in the synthesis of β-glucan were mapped to centromeric region of 7HL chromosome arm using the 4BS.7HL Asakaze-Manas translocation line (Cseh et al., 2011), and on chromosomes 1HS and 7H using Mv9kr1-Igri wheat-barley addition lines (Cseh et al., 2013). These genes are of particular interest for producing reinforced wheat varieties, as β-glucans are potential cholesterol-lowering polysaccharides (Kerckhoffs et al., 2003) and non-specific immune-activators (Allendorf et al., 2005). Moreover, β-glucans are present in higher proportions in barley than in

wheat grain. Finally, 7HL chromosome arm was shown to contribute to increased salt tolerance relative to wheat, both during germination and in the early developmental stages in Asakaze-Manas wheat-barley crosses (Turkosi et al., 2016); Cseh et al., 2013; Dulai et al., 2010; Darko et al., 2017; Molnár-Láng et al., 2013a).

Overall, barley is a rich source of favorable alleles for wheat improvement. At the same time, wheat-barley chromosome (-arm) additions represent an excellent model system for identifying and understanding the molecular basis of functional interactions between the host genome and alien chromatin. Unlike chromosome translocation lines, for example, they make it possible to study the impact of alien introgression on the functionality of the host genome without a possible interference from the insertion of alien chromatin into wheat chromosomes. These lines are also more stable as compared to hybrids and first generation of backcrosses, which suffer from a strong transcriptomic shock and genomic rearrangements after hybridization. Finally, wheat and barley are the two Triticeae species for which most genomic and genome sequence resources are available, allowing for in depth structural and functional genomics analysis.

2.2 Nuclear genomes of wheat and barley

2.2.1 Origin and evolution of wheat and barley

Wheat and barley belong to tribe Triticeae which diverged about 26 million years ago (MYA) from other members of Poaceae. Later, *Hordeum* diverged from other Triticeae about 13 MYA, and *Triticum* and *Aegilops* diverged from *Secale* about 11 MYA (Murat et al., 2014). According to the most widely recognised scenario, hexaploid wheat arose from two rounds of hybridization events between three ancestral diploid species of the Triticeae tribe (Figure 5). The first hybridisation, about 0.5-0.8 MYA, involved *T. urartu* Thum. ex Gandil. which provided the A genome, and an unknown species from the Sitopsis section that provided the B genome (closely related to *Ae. speltoides*), and resulted in tetraploid wild emmer *T. turgidum* L. ssp. *dicoccoides* (Körn. ex Asch. & Graebn) Thell. (AABB). The second hybridization occurred approximatively 0.2-0.4 MYA, between the tetraploid wheat (*T. turgidum* L. ssp. *Dicoccum* (Schrank. ex Schübl.)

Thell.) with the AABB genome and wild diploid goat grass (*Ae. tauschii* Coss.) that provided the D genome (Petersen et al., 2006) and produced the hexaploid wheat *T. aestivum* (AABBDD, 2*n*=6*x*=42). Both allopolyploid wheats (tetraploid and hexaploid) were domesticated ~10,000 YA and gave rise to tetraploid pasta wheat (*T. turgidum* ssp. *durum* Desf.) and the modern bread wheat (Dubcovsky and Dvorak, 2007).

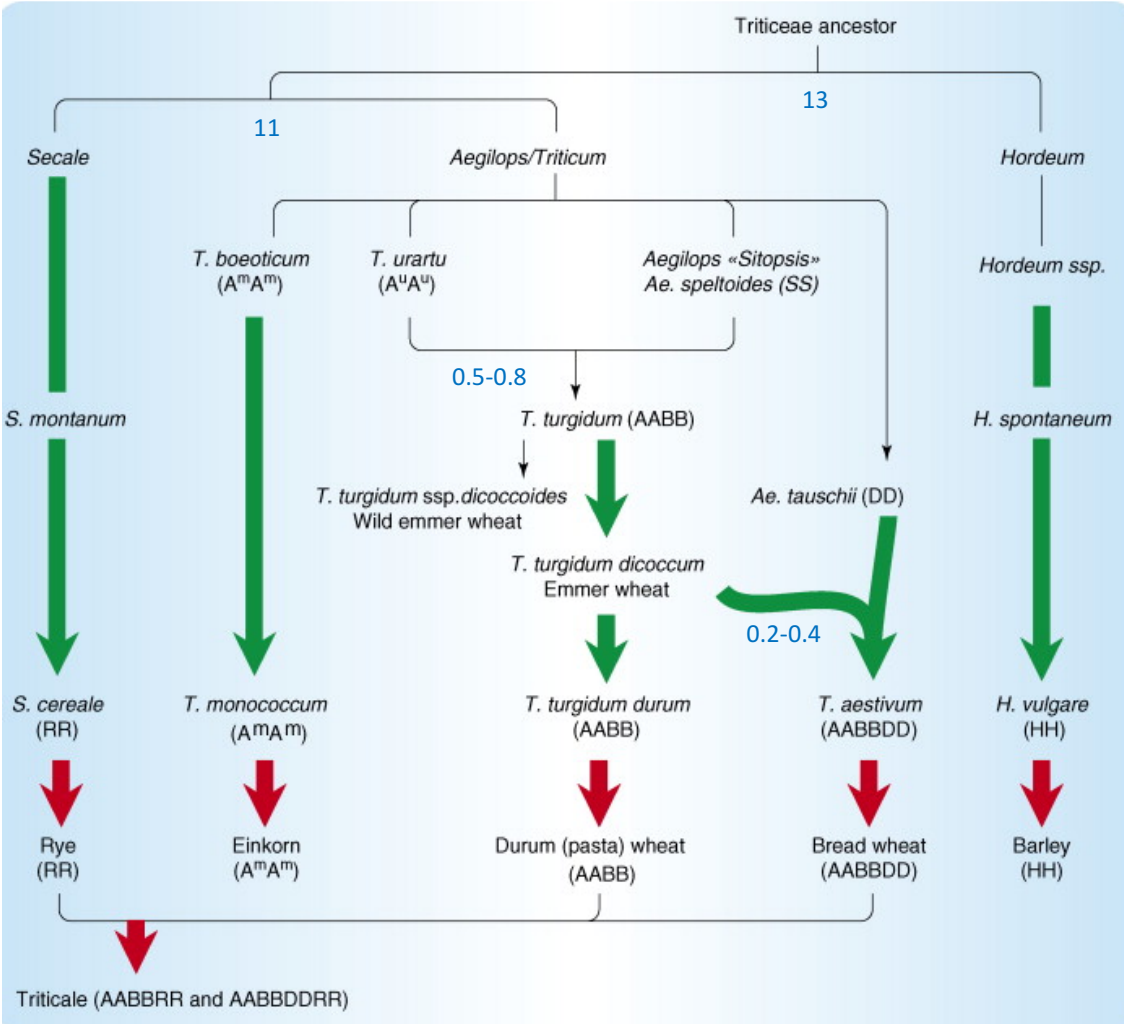


Figure 5: The evolution of the Triticeae species from a common ancestor - adapted from (Feuillet et al., 2008). The different evolutionary hybridization (black arrows), domestication (green arrows) and selection (red arrows) steps that have resulted in modern wheat, barley and rye

Recent studies based on genome assemblies of bread wheat and diploid relatives (Ling et al., 2013; Avni et al., 2017; Luo et al., 2017) provided novel insights into the phylogeny and evolutionary history of wheat. Notably, the time of divergence between the A and B genomes from a common ancestor was dated to about 7 MYA (Marcussen et al., 2014). Furthermore, the

D genome of hexaploid wheat is believed to originate from multiple rounds of hybridization events involving the progenitors of the A and B genomes without a change in ploidy level (homoploid speciation) between 5 and 6 MYA (Marcussen et al., 2014; Li et al., 2015; Sandve et al., 2015). The question of mono- or poly-phyletic origin of the progenitor of the B genome is still a matter of discussion. This genome could originate from a unique, ancient *Aegilops* species that remains unknown, or is extinct, but also from hybridization of several unknown parental *Aegilops* species from the Sitopsis section (Feldman and Levy, 2015). Tese novel findings expand our understanding of modern hexaploid wheat genome structure and composition.

Compared to wheat, barley has a straightforward taxonomy and evolution, with the term 'barley' referring to only one species *H. vulgare* (Nevo, 1992). Several different sub-species have been identified, including the cultivated subspecies *H. vulgare* ssp. *vulgare* and wild sub-species *H. vulgare* ssp. *spontaneum* (C. Koch) Thell. that can still be found growing in original habitats around the Fertile Crescent, but also in Aegean region of the Mediterranean, south eastern Iran, and central Asia, including Afghanistan and the Himalayan regions. The most widely accepted hypothesis is that cultivated barley *H. vulgare* ssp. *vulgare* was domesticated in the Fertile Crescent from *H. vulgare* ssp. *spontaneum* in one event dated approximately 10,000 years ago. However, an alternative hypothesis proposes that additional independent domestication events occurred in Iran (Saisho and Purugganan, 2007) and at the Horn of Africa (Orabi et al., 2007). The resolution of barley domestication is further complicated by gene flow between wild and cultivated species in the regions of sympatry (Mascher et al., 2016).

As a result of similar evolutionary trajectory, wheat and barley genomes share common genomic features. As for other members of Triticeae, wheat and barley are characterized by large genomes with high (>80%) repetitive DNA content. With ~16 Gb/1C, *T. aestivum* genome is 100 times the genome size of a model plant species *Arabidopsis thaliana* (L.) Heynh., while *H. vulgare* possess one of the largest genomes among diploid plant species (~5.3 Gb/1C). Due to relatively young evolutionary history of Triticeae, genome synteny is well conserved across the tribe and genomes of all diploid species (and sub-genomes in polyploids) are divided to seven chromosomes (basic chromosome number x = 7).

Large genome size and high DNA repeat content represent a huge challenge for sequencing and assembling the Triticeae genomes. Due to the importance of wheat and barley, it was very important to obtain a genome reference for both species in order to support research and breeding efforts. However, unlike maize and rice, which are also important cereal crops and were sequenced in 2009 and 2005, respectively (International Rice Genome Sequencing, 2005a; Schnable et al., 2009a), the reference genome sequences of wheat and barley were not available until very recently. After more than ten years of hard work, the International Wheat Genome Sequencing Consortium (IWGSC) and the International Barley Sequencing Consortium (IBSC) released recently gold standard reference sequences of the two species (Mascher et al., 2017; IWGSC, 2018). These resources provide novel insights into the content, structure and function of barley and wheat genomes and allow genome-wide studies at an unprecedented resolution.

2.2.2 Genome sequence resources for wheat and barley

2.2.2.1 Approaches for sequencing wheat and barley genomes

Wheat genome sequencing history started in 2005 with the launch of the International Wheat Genome Sequencing Consortium (IWGSC) whose aim was to produce a high-quality, physical map-based, ordered, and annotated genome sequence comparable in quality to the rice genome sequence. This initiative chose a 'chromosome-based' and BAC-by-BAC based physical map approach (Figure 6B and 6C). In order to reduce the genome complexity, the genome was dissected into a group of three chromosomes (1D, 4D, 6D), one chromosome (3B) and 34 arms of the remaining 17 chromosomes. IWGSC choose to sequence the hexaploid bread wheat genome of *T. aestivum* cv Chinese Spring for which a large set of genetic stocks is available, including telosomic lines that enabled flow sorting of individual chromosome arms (Sears, 1978). The chromosome-based approach facilitated the release of chromosome-based draft genome sequence (Chromosome Survey Sequences, CSS) (International Wheat Genome Sequencing, 2014), the pseudomolecule assembly of the largest (~1 Gb) wheat chromosome 3B (Choulet et al., 2014), and ultimately, the complete genome reference sequence IWGSC RefSeq_v1.0

(IWGSC, 2018) obtained by integrating the CSS with whole-genome shotgun Illumina sequencing assembled by the NRGene DeNovoMAGIC™ platform.

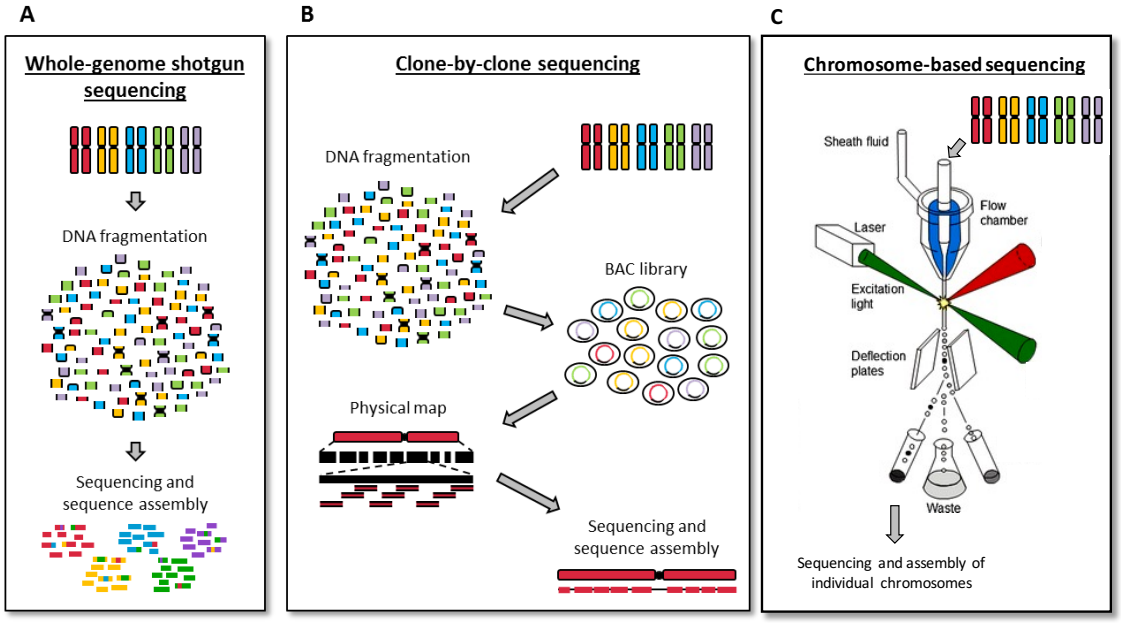


Figure 6: Different genome sequencing approaches: (A) whole-genome shotgun sequencing, (B) Clone-by-clone sequencing, (C) Chromosome-based sequencing – adapted from (Doležel et al., 2012; Choulet et al., 2014). (A) Whole-genome shotgun (WGS) sequencing is an approach subsuming fragmentation of the whole genome into small pieces of a defined length which are then sequenced. The whole set of data is then used for the sequence assembly. WGS sequencing represents straightforward, fast and relatively cheap procedure. (B) Clone-by-clone sequencing strategy is also known as hierarchical or BAC-by-BAC sequencing, the latter reflecting the fact that BAC vectors are currently most widely used for the clone-by-clone sequencing. The first step, which precedes the very sequencing, is the construction of a physical contig map. Physical contig map is represented by an ordered set of overlapping BAC clones assembled into contigs and covering the genome. The minimal set of overlapping clones ordered along the chromosomes and representing whole genome (minimal tilling path - MTP), is selected out of the whole physical map and sequenced. Individual MTP clones are fragmented before sequencing into small pieces and subsequently either cloned into a plasmid vector and sequenced using Sanger chemistry or sequenced directly by next-generation sequencing technologies, similarly as in the WGS approach. The physical contig map then provides a framework for merging of sequence contigs into scaffolds and finally for a construction of pseudomolecules of all individual chromosomes. (C) The chromosome-based approach, adopted by the IWGSC, begins with the construction of a BAC library from flow-sorted chromosomes/arms. All individual clones in the library are fingerprinted and a physical contig map is constructed based on overlaps between the fingerprints. Then, the minimal tilling path (MTP) is established and clones representing the MTP are shotgun sequenced using NGS technologies. Final steps of the sequencing are the sequence assembly, construction of a pseudomolecule and its validation.

In parallel, other teams attempted wheat genome sequencing through whole-genome sequencing approach solely (Figure 6A), with the aim to deliver whole-genome data in a short time and for a reasonable cost (Brenchley et al., 2012; Chapman et al., 2015; Zimin et al., 2017), for which the work of Benchley and colleagues was the first wheat whole genome draft sequence released. These studies provided rough insights into the wheat genome composition and genic content, with increased sequence contiguity due to the use of the improved sequencing technologies and assembly software. However, the assemblies were largely incomplete and fragmented, often lacking information on chromosome anchoring and/or annotation, thus limiting their use in this Ph.D. project.

Barley genome sequencing strategy was more straightforward, with the International Barley Sequencing Consortium (IBSC) launched in 2006 (Schulte et al., 2009). It envisaged a similar road map as IWGSC for wheat, with the exception of the chromosome-based strategy, which was not considered necessary for the assembly of the smaller diploid genome. IBSC sequenced *H. vulgare* ssp. *vulgare* cv Morex which is a six-row malting variety and which has been used frequently in many studies, including the identification of genes involved in malting quality traits such as malt extract, diastatic power, and grain protein and their regulation (Kleinhofs et al., 1993).

However, the first whole genome sequence of barley was obtained for cv Betzes after sequencing flow-sorted chromosomes (-arms) providing a virtual order of barley genes (genome-zipper) (Mayer et al., 2011a). IBSC released the first draft genome assembly in 2012 (International Barley Genome Sequencing et al.), followed by a complete reference genome assembly in 2017 (Mascher et al.). Whole genome shotgun sequencing was also done in four barley cultivars Bowman, Barke, Igri and Haruna Nijo, as well in *H. vulgare* ssp. *Spontaneum*, the progenitor of cultivated barley (International Barley Genome Sequencing et al., 2012).

As a result of different efforts to sequence wheat and barley genomes, various genomic resources were gradually released during the course of this doctorate study (Figure 7). Each of them

provided a different type of information and affected the strategy of our analysis. However, our main analysis was done using the reference genomes of wheat and barley and the two resources are described in the following.

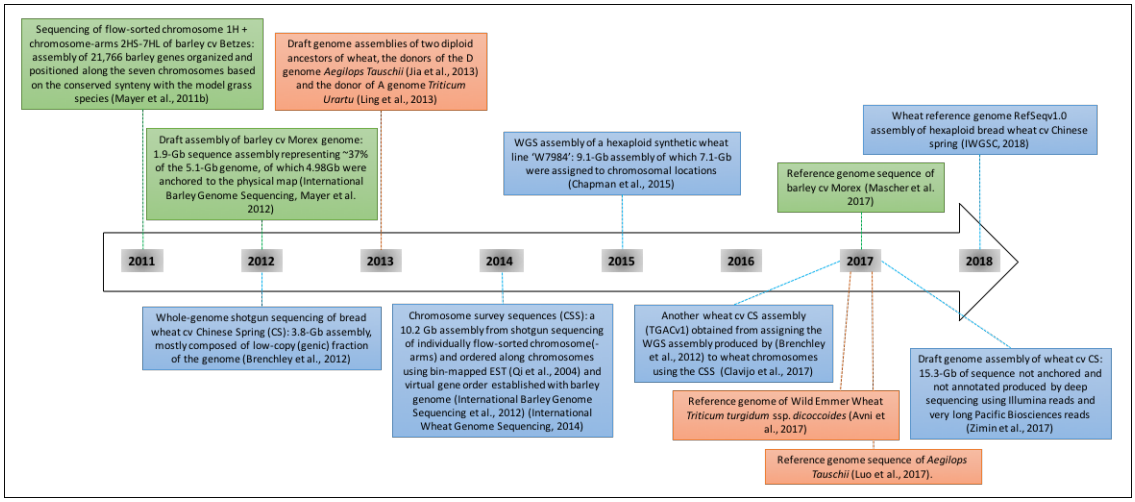


Figure 7: The different genome sequences released for wheat (blue), barley (green) and diploid progenitors of wheat (orange).

2.2.2.2 Reference genome sequences of wheat and barley

In 2017, Mascher and colleagues published the first reference genome sequence of diploid barley *H. vulgare* ssp. *vulgare* cv Morex. This assembly was produced using a hierarchical approach (Mascher et al., 2017). First, BAC clones selected from the minimal set of overlapping clones covering chromosomes called Minimal Tilling Path (MTP) were sequenced by Illumina technology. Individual BAC clone sequences were then assembled into contigs and merged into scaffolds based on overlaps between adjacent clones in the physical-map. The assembly of scaffolds was then ordered into super scaffolds using genetic-linkage and optical maps, and the super-scaffolds were assigned to chromosomes based on genetic map constructed after genetic-population sequencing (POPSEQ). Finally, BAC-based super-scaffolds were ordered and oriented along chromosomes based on three-dimensional proximity information obtained by chromosome conformation capture sequencing (Hi-C) analysis. The chromosome-scale assembly (pseudomolecules) of the barley genome consists of 6,347 super-scaffolds ordered along the

seven chromosomes of barley and represents 4.79 Gb (~ 95%) of the 5.1Gb estimated genome size. Of these, 4.54 Gb were assigned to precise chromosomal locations. The annotation revealed 83,105 putative gene loci including protein-coding genes, non-coding RNAs, pseudogenes and transcribed transposons. These loci were filtered and divided into 39,734 high-confidence genes and 41,949 low-confidence genes on the basis of sequence homology to related species. Overall, about 1.4% of the assembly was annotated as coding sequences, and 80.8% were annotated as transposable elements.

In 2018, the reference assembly of hexaploid wheat genome *T. aestivum* cv Chinese Spring will be published by IWGSC (IWGSC, 2018). This assembly was obtained by integrating a draft whole genome *de novo* assembly obtained by Illumina sequencing and assembled using the DeNovoMAGIC2 assembly algorithm (Baruch et al., 2016), together with the other genome sequencing resources which have been produced by the wheat community so far. The whole genome assembly was assigned to individual chromosome arms by alignment to the chromosome survey sequences (CSS), and sequence scaffolds were ordered along chromosomes using Hi-C analysis and high density genetic map (POPSEQ). Scaffolds were arranged to super-scaffolds using a combination of physical maps, chromosome-BAC library sequencing and optical maps. The resulting assembly is a fully annotated and anchored reference sequence, representing ~94% of the wheat genome organized in 21 pseudomolecules that compose the A, B, and D sub-genomes of hexaploid bread. The annotation identified 107,891 high confidence protein coding loci, with relatively equal distribution across the A, B, and D sub-genomes (35,345, 35,643, and 34,212, respectively) out of which 82.1% had a function assigned. In addition, 161,537 other protein coding loci were classified as low confidence genes, representing partially supported gene models, gene fragments, and orphans. A total of 3,968,974 transposable elements (TEs) copies were identified belonging to 505 families and representing 85% of the genome, with very similar proportions on the three sub-genomes (A: 86%, B: 85%, and D: 83%).

2.2.3 The genomes of wheat and barley

The release of wheat and barley reference genome sequences (Mascher et al., 2017; IWGSC, 2018) provides unprecedented support for genome-scale analyses, and allows in-depth analysis of their genome structure, composition and dynamics. Here, we describe these novel resources and summarize the current knowledge of wheat and barley genomes

2.2.3.1 Wheat and barley genome organization

In plants, it is known that gene content is quite conserved across species, or at least does not vary proportionally with the genome size (Bennetzen, 2000). This implies that the distribution of genic space (part of the DNA encoding transcriptionally active genes) is non-random, but rather structured relative to genome size (International Rice Genome Sequencing, 2005b; Paterson et al., 2009; Schnable et al., 2009b; International Brachypodium, 2010). The tendency is a uniform distribution of genes along chromosomes in genomes which size is smaller than 500Mb such as rice (389Mb), while large genomes such as maize (2.3GB), sorghum (730Mb), barley (5.2Gb) and wheat (16Gb) have an increasing gene density towards the telomeres. Moreover, it has been shown in wheat that the intergenic distances are extremely variable (average 104 ± 190 kb) and that a majority (about 73%) of genes are organized in small islands (Rustenholz et al., 2011; Choulet et al., 2014), a feature that would be shared among grass genomes (Gottlieb et al., 2013). The analysis of distribution of structural and functional features (gene and TE density, recombination rate, expression breadth, …) confirmed the structural and functional partitioning into five chromosome compartments for all 21 wheat chromosomes and all 7 barley chromosomes (Figure 8): the two chromosome extremities (distals) which exhibit high recombination rates, with high gene content and low TE-density, and are also enriched in genes expressed in fewer conditions and genes involved in defense response and reproductive processes; the two interstitial regions that are characterized by increased TE density, as well as intermediate gene density and recombination rate; and the centromeric and peri-centromeric region (proximal) which form the last chromosomal compartment characterized by the absence of recombination, low gene density

and high density of TE, and which contain more genes related to housekeeping processes, such as photosynthesis and respiration and show a high level of expression breadth (Choulet et al., 2014; Pingault et al., 2015; Mascher et al., 2017).

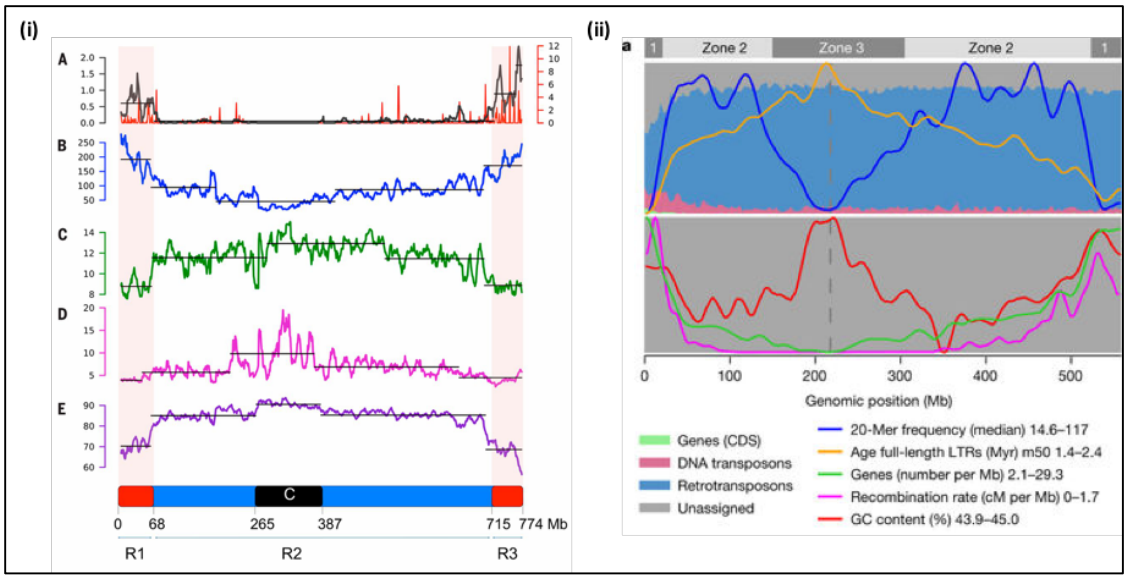


*Figure 8: Structural and functional partitioning of (i) wheat 3B and (ii) barley chromosomes – adapted from (Choulet et al., 2014; Mascher et al., 2017). On the left (i) panel is represented the distribution and segmentation analysis of (**A**) meiotic recombination rate (cM/Mb in sliding window of 10 Mb in black and 1 Mb in red); (**B**) gene density (CDS/10 Mb); (**C**) expression breadth; (**D**) average number of alternative spliced transcripts per expressed gene; and (**E**) TE content along the 3B pseudomolecule. Distal regions of the chromosome R1 and R3 are represented in red. In (C), the centromeric/pericentromeric region is in black.*

2.2.3.2 Synteny and collinearity

The genome of bread wheat (*T. aestivum*) is composed of three homeologous sub-genomes (A, B and D) of similar size (~5.5Gb), each carrying highly similar gene set with an average identity of ~97% between homeologous gene copies (Choulet et al., 2014). Using a set of 39,474 wheat homeologous gene groups, comprising 114,495 genes, it was shown that overall synteny between the seven triplets of homeologous chromosomes was high, indicating that no massive rearrangements occurred since the A, B, and D genomes diverged ~5 MYA. However, collinearity between homeologs was disturbed by inversions occurring on average every 74.8 Mb involving blocks of ten or more genes. Among the 114,495 homeologous genes, 80% were found organized in macro-synteny (still present at their ancestral position), and 70% were organized in

collinear blocks (intervals with a highly-conserved gene order). Moreover, the level of synteny was variable along the chromosomes and this variation correlated with the functional partitioning of chromosomes, with a higher proportion of syntenic genes found in the interstitial compared to the distal telomeric and proximal regions, respectively. Overall, at the whole genome level, distal chromosomal regions are not only the preferential targets of meiotic recombination, but are also the fastest evolving chromosome regions.

Comparative genome analysis between wheat and barley reported a high degree of similarity between the two species with the exception of wheat chromosome 4A, which is characterized by a large-scale inversion and two interchromosomal translocations (Miftahudin et al.; Mickelson-Young et al., 1995; Nelson et al., 1995; Mayer et al., 2011b). On a structural basis, none of the individual wheat A, B, or D subgenomes was more closely or distantly related to the barley genome with numerous variations apparent in only one or two wheat subgenomes (Mayer et al., 2011b). However, it was evidenced that some regions that appear to be present in barley were lacking counterparts in any of the homeologous wheat chromosomes (e.g., 1AS, 1AL, 2AL, and 2DL, all long arms of homeologous group 5 chromosomes) (Mayer et al., 2011b). High similarity between wheat and barley genomes reflects the high degree of conservation between grass species over 50–70 MY of divergent evolution (Bolot et al., 2009).

2.2.3.3 Gene loss in homeologous wheat genomes

Gene loss in a polyploid species may generate patterns of a biased genome fractionation, i.e. sub-genome dominance, as suggested for maize (Woodhouse et al., 2010) and other grasses (Schnable et al., 2012). Wheat homeologous genes are present in different copy numbers, with at least one copy on two different sub-genomes. The proportion of genes in homeologous groups was shown to be highly similar in the three sub-genomes: 74% (A), 72% (B), and 77% (D). Although a majority of genes is present in homeologous groups, only 55% of the groups contained only one single gene copy per sub-genome (1:1:1 configuration) (IWGSC, 2018). A comparison of gene sequences of hexaploid bread wheat with those in putative donors of the A, B and D wheat

genomes indicated a limited gene loss during the evolution of the hexaploid wheat genome and frequent gene duplication after the sub-genomes formed the bread wheat genome (Eversole et al., 2014a). Detailed analysis facilitated by the reference genome assembly showed that the loss of individual homeologs occurred at very similar levels in the A, B, and D genomes and affected 10.6% (0:1:1), 10.2% (1:0:1), and 9.4% (1:1:0) of the homeologous groups, respectively (IWGSC, 2018). These data did not provide support for biased fractionation of sub-genomes in bread wheat. Instead, a gradual gene loss and gene movement among the three sub-genomes was observed that is proportional with the time span since the hybridization events occurred to form the bread wheat genome (IWGSC, 2018).

2.2.3.4 Transposable element content and niches

If gene content is quite conserved across plant species, the content, composition and genomic distribution of repeat elements is highly variable and generally reflects genome size (Heslop-Harrison and Schwarzacher, 2011). Large plant genomes consist mainly of highly similar copies of repetitive elements such as long terminal repeat (LTR) retrotransposons and DNA transposons, which play a major role in genome size expansion (Bennetzen, 2002). Detailed analysis of the wheat genome repeat content showed that most of the repetitive elements (64%) are class I repetitive elements, represented mostly by LTR retrotransposons Gypsy (45%) and Copia (16%) (IWGSC, 2018). About 15% of the genome sequence is composed of class II elements (DNA transposons), most of them CACTA elements (15%). Transposable elements are distributed along the whole chromosome length with the highest density in centromeric-pericentromeric regions (IWGSC, 2018). DNA transposons Mutator, Harbinger and MITEs were found close to genes, whereas LTR retrotransposons and CACTA elements locate at larger distances from centromere (Choulet et al., 2014). The repeat analysis revealed balanced proportions of TEs between sub-genomes with the different TE families maintained in similar proportions independently in the three diploid lineages. Except for small variations in the global proportion (~80%) of transposable

elements, barley exhibits very similar composition and distribution of repetitive DNA elements as wheat (Mascher et al., 2017).

2.2.3.5 Genic content and gene expression

Recent RNA-seq analysis of wheat provided new insights into gene transcription at genome and sub-genome level. The analysis of 850 mRNA samples, derived from 32 tissues at different developmental stages and/or subjected to different stress treatments mapped to the wheat reference sequence annotation, revealed that 31.0 % of wheat genes are expressed in over 90% of tissues and about 21.5% of genes are expressed in 10% or fewer tissues (IWGSC, 2018). About 8,252 HC genes have a tissue-specific expression profile, out of which about half are the genes expressed in reproductive tissues (microspores, anther and stigma/ovary) (IWGSC, 2018). Differential expression between samples was shown to be primarily driven by tissue, rather than growth stage or stress factors (IWGSC, 2018). It was also shown that homeologous genes tend to have similar expression patterns and that more divergent expression is observed for homeologs in non-syntenic chromosome context than homeologous genes in syntenic chromosome positions (with similar expression patterns) (IWGSC, 2018).

Although an RNA-seq experiment involving a combination of 16 different tissues, each with three biological replicates, was performed for the annotation of the barley cv Morex reference genome (Mascher et al., 2017), no transcriptome profiling across tissues and developmental stages was performed. However, based on previous transcriptome analysis of the draft genome, about 72–84% of HC genes were expressed in all mRNA samples, among which 36–55% were differentially regulated between samples, and about 55% had evidence of alternative splicing (International Barley Genome Sequencing et al., 2012).

2.3 Introduction to transcriptomic analysis

2.3.1 Molecular definitions

In the framework of this Thesis, we performed transcriptomic analysis of a wheat-barley chromosome-arm addition line using RNA-sequencing method. Producing and analysing the transcriptomic data itself represented an important part of this work as it necessitated deep understanding of the bioinformatics and statistics aspects of RNA-seq technology, and conditioned the quality of the results of our study. Therefore, in this section, general concepts of transcriptomic analysis that impacted on our strategy for the analysis of the wheat-barley addition line are introduced.

2.3.1.1 The transcriptome

Transcriptomics is the field of study of the 'transcriptome', which is defined as the complete set of RNA transcripts that are produced by the genome of an organism at any one time (Wang et al., 2009a). In contrast with the genome, which is characterized by its general stability, transcriptome changes across space and time. In fact, an organism's transcriptome varies depending on many factors, including developmental stage, tissue type, organ, and environmental conditions. Measuring gene expression in different tissues, conditions, or time points gives information on how the genes are regulated and reveals details of an organism's biology. It can also help to infer the functions of previously unannotated genes. Transcriptomics typically analyses populations of messenger RNA molecules (mRNA) as they represent intermediates between genes and proteins. Studying the full range of transcripts produced by an organism necessitates the use of high-throughput methods and the best contemporary techniques are various forms of microarray technologies, which quantify a set of predetermined sequences, and RNA sequencing (RNA-seq), which uses high-throughput sequencing to capture all transcripts (Lowe et al., 2017a). Ultimately, the comparison of transcriptomes allows the identification of genes that are differentially expressed in distinct cell populations, or in response to different stimuli.

2.3.1.2 Gene transcription

In eukaryotes, a majority of genes are characterized by intron/exon structure that produces molecules of RNA, which can either be non-coding RNA (ncRNA) with a direct function, or a messenger RNA (mRNA) and encode a protein. Coding genes are composed of several features containing information that allows the gene to express its functionality (Figure 9). These features include open reading frame (ORF), the part of a gene sequence that has the potential to be translated into amino acids; regulatory sequences located at one or at both extremities of the gene such as promotors, enhancers and silencers that regulate the transcription; regulatory sequences of the translation, among which 5' and 3' untranslated regions (UTR). The ORF and UTRs represent the transcribed region of the gene and include the coding regions called exon, and non-coding regions called introns (Figure 9). However, not all genes encode proteins, as for example genes producing long non-coding RNA (lncRNA), which can be transcribed, but lack ORF and therefore are not translated into proteins (Kung et al., 2013), or ribosomal RNA (rRNA) genes which encode for the RNA component of ribosomes and are essential for protein synthesis in all living organisms.

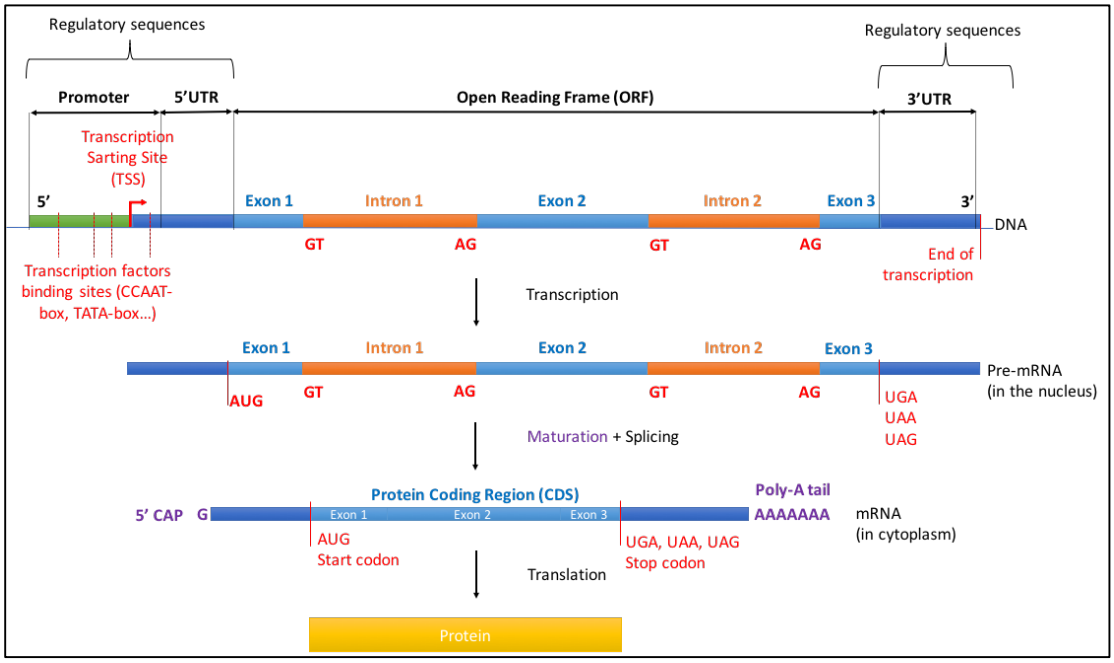


Figure 9: Structure of eukaryotic protein coding genes and transcription – adapted from (Alberts B, 2002).

The transcription is a process by which the information carried by a strand of DNA is copied into a molecule of RNA by RNA polymerase (i.e. RNA polymerase II in the case of transcription of protein coding genes) to obtain the so called "transcript" (Kornberg, 2007). It is the first step of the expression of genes into proteins, and occurs primarily in the nucleus of eukaryotes organisms. The first transcript immediately produced after transcription of a gene, called precursor mRNA transcript (pre-mRNA), is not directly translated into a protein but has to go through a maturation step: first, a '5' cap' is formed by the addition of a modified guanosine (7-methylguanosine residue) at 5' end, then, a poly(A)-tail is added at the 3' end of coding transcripts as well as to some long non-coding RNAs during a step called polyadenylation. The 5' cap and 3' poly(A) tail allow for better stability of the RNA molecules which can then be transported to cytoplasm. Further to these steps of maturation, pre-mRNA is modified by the excision of introns through a process called splicing (Kruger et al., 1982).

However, a particular combination of exons and introns may be retained or spliced out during a process called alternative splicing (Kruger et al., 1982; Quelle et al., 1995). Because of the alternative splicing, the same gene can produce several transcripts, called transcript variants or isoforms, which differ by the combination of introns and exons, and thus potentially result in synthesis of several different proteins. Alternative splicing increases the diversity of an organism's transcriptome and proteome (Shang et al., 2017). In plants, alternative splicing has been shown to play an important role in the regulation of flowering time (Lee et al., 2013; Pose et al., 2013), circadian clock (Filichkin et al., 2015), environmental stress response (Egawa et al., 2006) and resistance to biotic stress (Staiger and Brown, 2013).

2.3.1.3 The evolving concept of transcriptome

The transcriptome such as described above, is a protein-centric view of molecular biology inherited from the concept of RNA being an intermediate molecule between DNA and proteins, defined by Crick (1958) as the 'central dogma'. However, large-scale transcriptomic studies have challenged this view and expanded the definition and role of the transcriptome. Various

microarray and RNA-seq technologies permitted large-scale characterization of various non-protein coding RNA and novel mechanisms of regulation of gene expression such as:

- Small regulatory RNA including microRNA (miRNA) (Lee et al., 1993; Reinhart et al., 2000), and small interfering RNA (siRNA) (Hamilton and Baulcombe, 1999), involved in post-transcriptional regulation of gene expression through RNA interference (RNAi) (Fire et al., 1998; Waterhouse et al., 1998). These discoveries demonstrated that RNA is not only an intermediate between DNA and protein, but may play a role in complex regulatory mechanisms
- RNA *trans*-splicing, evidencing that one protein can result from the combined information encoded in multiple transcripts (Borst, 1986; Takahara et al., 2000)
- RNA editing, describing the enzymatic modification of RNA (Eisen, 1988) and demonstrating that the information in the RNA sequence is not completely reflecting DNA sequence
- Long non-coding RNAs (lncRNA), which are long transcripts (200nt or more) that do not code for proteins, and are located either intergenic, intronic and antisense to protein-coding gene sequences (Carninci et al., 2005; Cheng et al., 2005; Kapranov et al., 2005). lncRNA were shown to play a role in the regulation of gene expression through epigenetic, transcriptional and post-transcriptional mechanisms (Cao, 2014).

As the knowledge of transcriptome is expanding, both gene and RNA definitions have evolved. In 2007, (Gerstein et al.) emphasized that "What has not changed is that genotype determines phenotype, and at the molecular level, this means that DNA sequences, determine the sequences of functional molecules. In the simplest case, one DNA sequence still codes for one protein or RNA. But in the most general case, we can have genes consisting of sequence modules that combine in multiple ways to generate products". The authors also propose a revised definition of what is a gene which can be resumed as "a union of genomic sequences encoding a coherent set of potentially overlapping functional products".

2.3.2 Techniques to study the transcriptome

The field of transcriptomics expanded with the development of sequencing technologies and with the development of techniques and approaches for large-scale bioinformatics analysis of transcriptomes (Figure 10).

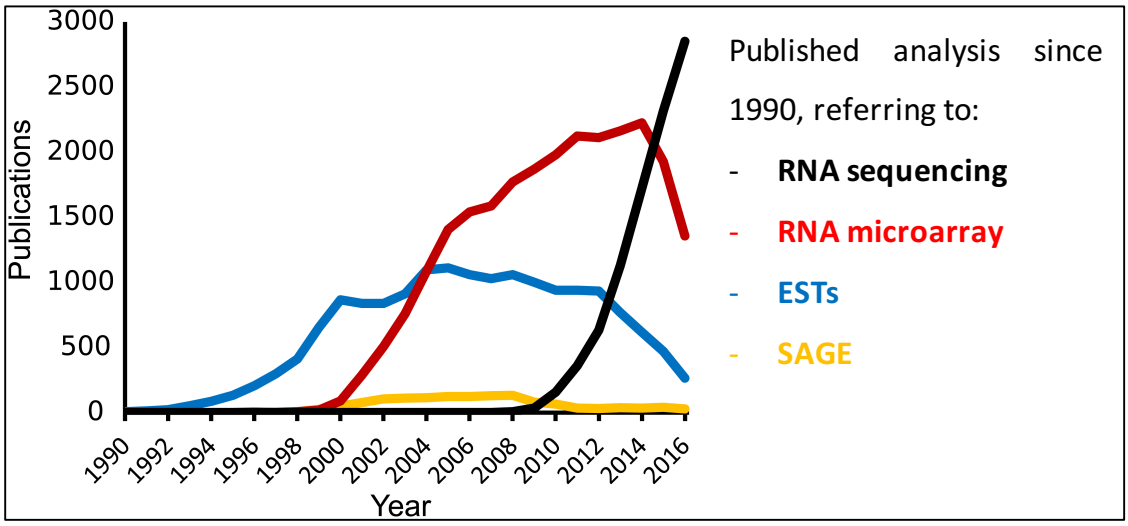


Figure 10: The use of transcriptomic methods over time – adapted from (Lowe et al., 2017a).

2.3.2.1 Expressed sequenced tags (ESTs)

The first attempts to study the whole transcriptome began in the early 1990s and involved the Sanger sequencing of cDNA molecules. These experiments produced so called Expressed Sequence Tags (ESTs) (Adams et al., 1991). ESTs came to prominence during the 1990s as an efficient method to determine gene content of an organism without sequencing the entire genome (Marra et al., 1998). However, it was a laborious approach and did not allow for high-throughput transcript quantification. It relied on Northern blotting (Alwine et al., 1977), hybridization on nylon membrane array (Cox, 2001), and reverse transcription polymerase chain reaction (RT-qPCR) (Bustin, 2000), and was limited to a small subset of the whole transcriptome. However, some of these quantitative techniques are still in use for characterization of the expression of one or more genes, and are useful in the validation of results obtained by less accurate high-throughput methods of transcriptome analysis (Lowe et al., 2017a).

2.3.2.2 Serial analysis of gene expression (SAGE)

In order to increase the amount of tags generated and also to estimate transcripts abundance, Serial Analysis of Gene Expression (SAGE) was developed as an extension of ESTs technology. In SAGE, cDNAs are digested into 11 bp "tag" fragments which are then concatenated and sequenced. The abundance of transcripts is directly quantified from the abundance of the tag which is representing it. Although SAGE method produces information on more genes than was possible when sequencing single ESTs, the sample preparation and data analysis are more labour intensive and only a portion of each transcript is analyzed. Moreover, isoforms are generally indistinguishable from each other and a significant portion of the short tags cannot be uniquely mapped to reference genome. These disadvantages strongly limited its use (Wang et al., 2009a). The true transcriptomic revolution came with the introduction of two major techniques: microarrays and RNA-sequencing, which became the predominant technologies in the contemporary studies as they allow large-scale characterization of transcriptomes.

2.3.2.3 Microarrays

Microarrays are hybridization-based methods which consist of measuring the abundance of a defined set of fluorescently labelled transcripts (cDNA) *via* their hybridization to an array of complementary probes. The first microarray study was published in 1995 (Schena et al.) and soon high-density arrays became the method of choice for transcriptional profiling (Figure10). Microarray technology allowed to assay thousands of transcripts simultaneously at reduced cost per gene and was less laborious as compared to older methods (Heller, 2002). However, this approach also suffers from some limitations. First, it requires a knowledge of genome sequence in order to design oligo probes. This makes it difficult to use the array technologies in non-model species. The second issue with microarrays is the high level of cross-hybridization and a limited dynamic range of detection resulting from background noise and saturation of signals (Lowe et al., 2017a). Finally, comparing expression levels across different experiments is often difficult and requires complicated normalization methods (Wang et al., 2009a).

2.3.2.4 RNA-sequencing

RNA-seq approach profits from the rapid progress in the Next Generation Sequencing (NGS) technologies (Lister et al., 2008; Mortazavi et al., 2008; Nagalakshmi et al., 2008). RNA-seq provides access to transcript sequences with nucleotide resolution and without prior knowledge of genic sequences. This potentially allows to access entire set of transcripts and represents substantial improvement over the microarrays (Table 1). The quantification of transcripts is relative to the number of reads obtained and is therefore more precise than quantification with microarrays. RNA-seq is currently the most popular method for large-scale transcriptome analysis allowing both identification of novel transcripts and their quantification. As it does not require prior sequence knowledge, and it is applicable to non-model species for which no reference genome is available (Wang et al., 2009a).

Table 1: Comparison RNA-seq versus Microarrays technologies – adapted from (Wang et al., 2009b; Lowe et al., 2017a).

Method	RNA-seq	Microarray
Principle	High-throughput sequencing	Hybridization
Throughput	High	Higher
Input RNA amount	Low ~1ng total RNA	High ~1µg mRNA
Labour intensity	High (sample preparation and data analysis)	Low
Prior sequence knowledge	None required, though genome sequence is useful	Reference transcripts requires for probes
Quantification accuracy	~90% (limited by sequence coverage)	>90% (limited by fluorescence detection accuracy)
Sequence resolution	Single base, can detect SNPs and splice variants (limited by sequence accuracy)	From several to 100 bp, can detect splice variants (limited by probe design and cross-hybridization)
Sensitivity	10^{-6} (limited by sequence coverage)	10^{-3} (limited by fluorescence detection)
Dynamic range	$>10^{-5}$ (limited by sequence coverage)	$10^{-3} – 10^{-4}$ (limited by fluorescence saturation)
Technical reproducibility	>99%	>99%

A typical RNA-seq experiment consists of two main steps (Figure 11). The first of them is a wet-lab experiment and involves RNA isolation, conversion of RNA transcripts into a library of

cDNA fragments adapted to sequencing by NGS technologies, and sequencing these fragments. Various protocols have been developed for sequencing different types of RNA, including mRNA, snRNA, lncRNA, and lowly expressed RNA, for various sequencing technologies (Table 2), and for a wide range of applications (i.e. quantification, annotation of novel transcripts, splicing variants analysis, single cell transcriptomics) (Hrdlickova et al., 2017).

Table 2: Sequencing technology platforms commonly used for RNA-Seq – adapted from (Lowe et al., 2017b).

Platform (Manufacturer)	Commercial release	Typical read length	Maximum throughput per run	Single read accuracy	RNA-Seq runs deposited in the NCBI SRA (Oct 2016) [74]
454 (Roche, Basel, Switzerland)	2005	700 bp	0.7 Gbp	99.9%	3548
Illumina (Illumina, San Diego, CA, USA)	2006	50–300 bp	900 Gbp	99.9%	362903
SOLiD (Thermo Fisher Scientific, Waltham, MA, USA)	2008	50 bp	320 Gbp	99.9%	7032
Ion Torrent (Thermo Fisher Scientific, Waltham, MA, USA)	2010	400 bp	30 Gbp	98%	1953
PacBio (Pacbio, Menlo Park, CA, USA)	2011	10,000 bp	2 Gbp	87%	160

NCBI, National Center for Biotechnology Information; SRA, Sequence Read Archive; RNA-Seq, RNA sequencing.

https://doi.org/10.1371/journal.pcbi.1005457.t002

The second step involves *in silico* analysis of RNA-seq data. The RNA-seq experiments generate large volumes of raw sequence reads, which have to be processed to obtain transcript sequences from the fragmented information. The process of RNA-seq analysis can be split into the following steps: reconstruction of transcript sequences from the sequencing reads, estimation of transcript abundance, and the comparison of transcript abundances in different conditions (tissues, developmental stages, physiological treatments). Actual RNA-seq data analysis depends on the aim of the analysis and usually requires a combination of various bioinformatics tools, depending on the experimental design, genomic and transcriptomic sequence resources available. In the next section, we will introduce different methods and approaches currently available for the analysis of RNA-seq data.

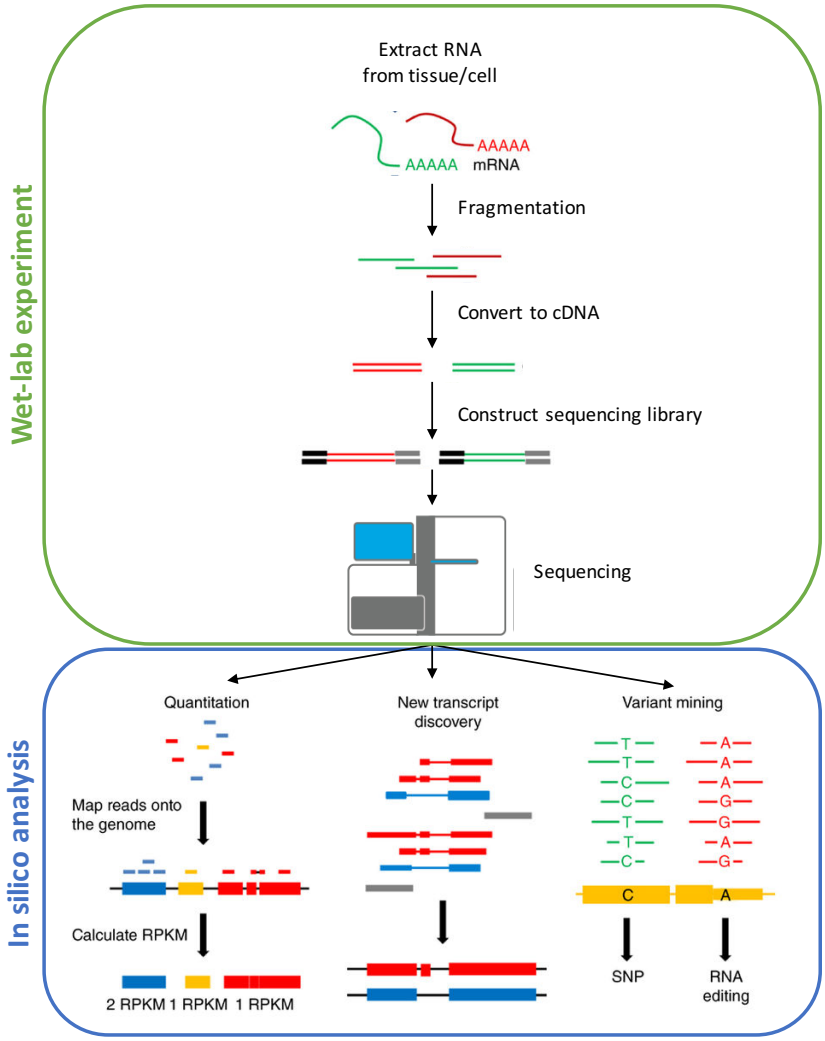


Figure 11: A typical RNA-seq workflow – adapted from (Zeng and Mortazavi, 2012).

Third generation sequencing technologies are becoming available since 2010 and provide long DNA sequence reads of a mean size ranging between tens to hundreds kilobases (Table 2). The two commercially available are Single Molecule Real-Time (SMRT) from Pacific Bioscience (PacBio) and nanopore sequencing (MinION) from Oxford Nanopore Technologies. Unlike the previous generation of NGS, the third-generation sequencing technologies do not involve DNA amplification before sequencing, which reduces sequencing errors. Together with long reads this makes of them excellent choice for novel transcripts annotation as well as isoform detection. However, lower throughput of current models prevents their use for transcript quantification and variant calling, where the accuracy depends on the quantity of reads produced (Lowe et al., 2017b). While a typical third generation sequencing run (e.g. PacBio RSII) yields ~100,000 reads,

dozens of million reads are obtained during a typical NGS sequencing run. The third-generation sequencing technologies are thus not yet the platform of choice for transcript quantification, but might change the way RNA-seq is performed in future.

2.3.3 Introduction to RNA-seq analysis

As mentioned above, RNA-seq analysis involves the reconstruction of a comprehensive transcriptome from short RNA-seq reads. After this step, the abundance of transcripts can be quantified and, if needed, the data are used to compare transcript abundance in different samples. The choice of a strategy for the reconstruction of transcripts depends on the availability of a genome, or transcriptome reference. The methods for transcript abundance quantification are selected based on the experimental design and the aim of the study (gene/transcript quantification, splicing variant analysis, SNP variant analysis, etc.).

2.3.3.1 Approaches for transcripts reconstruction

Any RNA-seq analysis starts by inferring which transcripts are expressed in the samples. There are three main strategies suitable for three different scenarios: a comprehensive reference transcriptome is available, a reference genome is available, no prior genic information is available (Figure 12).

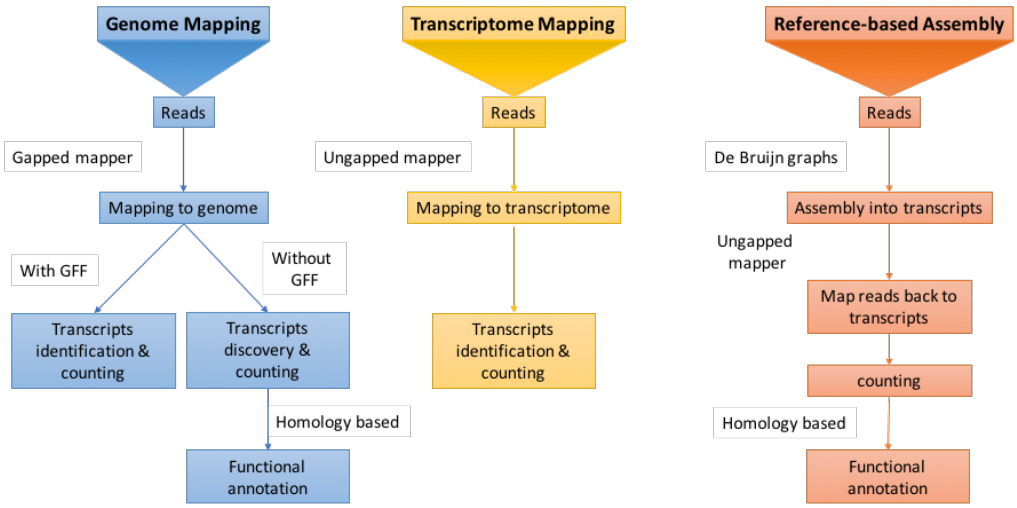


Figure 12: Read mapping and transcript identification strategies - adapted from (Conesa et al., 2016).

2.3.3.1.1 Transcriptome mapping

If a reference transcriptome (either all transcript sequences from an annotated genome, a *de novo* transcripts assembly, or a library of full-length cDNA) is available, the transcript composition can be inferred by mapping RNA-seq reads directly onto the reference transcriptome. As RNA-seq reads are obtained by transcript sequencing, the alignment of reads against the existing transcript sequences should not allow any gap in the alignment (meaning that the read sequence cannot be split in order for different parts to map non-contiguous transcripts sequences), and classical (unspliced) mappers such as Bowtie (Langmead et al., 2009), Bowtie2 (Langmead and Salzberg, 2012) and BWA (Li and Durbin, 2009) can be used.

Important parameters to consider during transcript mapping are the strandedness of RNA-seq library (e.g., in stranded library protocols, the information of the DNA strand origin from which the RNA was produced is kept), the number of mismatches to allow in the alignment of the read with the transcript sequence, the length of sequencing reads, the type of reads - single-end (SE) or paired-end (PE) – depending if the cDNA fragment is sequenced from one or both ends respectively, and the length of sequenced fragments. Mapping to the transcriptome is generally faster, but does not allow for *de novo* transcript discovery, though it is the most straightforward RNA-seq analysis pipeline possible, where transcript identification and quantification occur simultaneously.

2.3.3.1.2 Genome-guided

When a reference genome is available, the transcriptome assembly can be built on it, including the assembly of previously annotated and novel transcripts. This strategy known as 'reference-based' or '*ab-initio*' assembly is performed in three steps: the splice-mapping of reads onto the genome, the clustering of overlapping reads at each locus into a graph of all possible isoforms, and traversing the graph to resolve the actual isoforms (Figure 13).

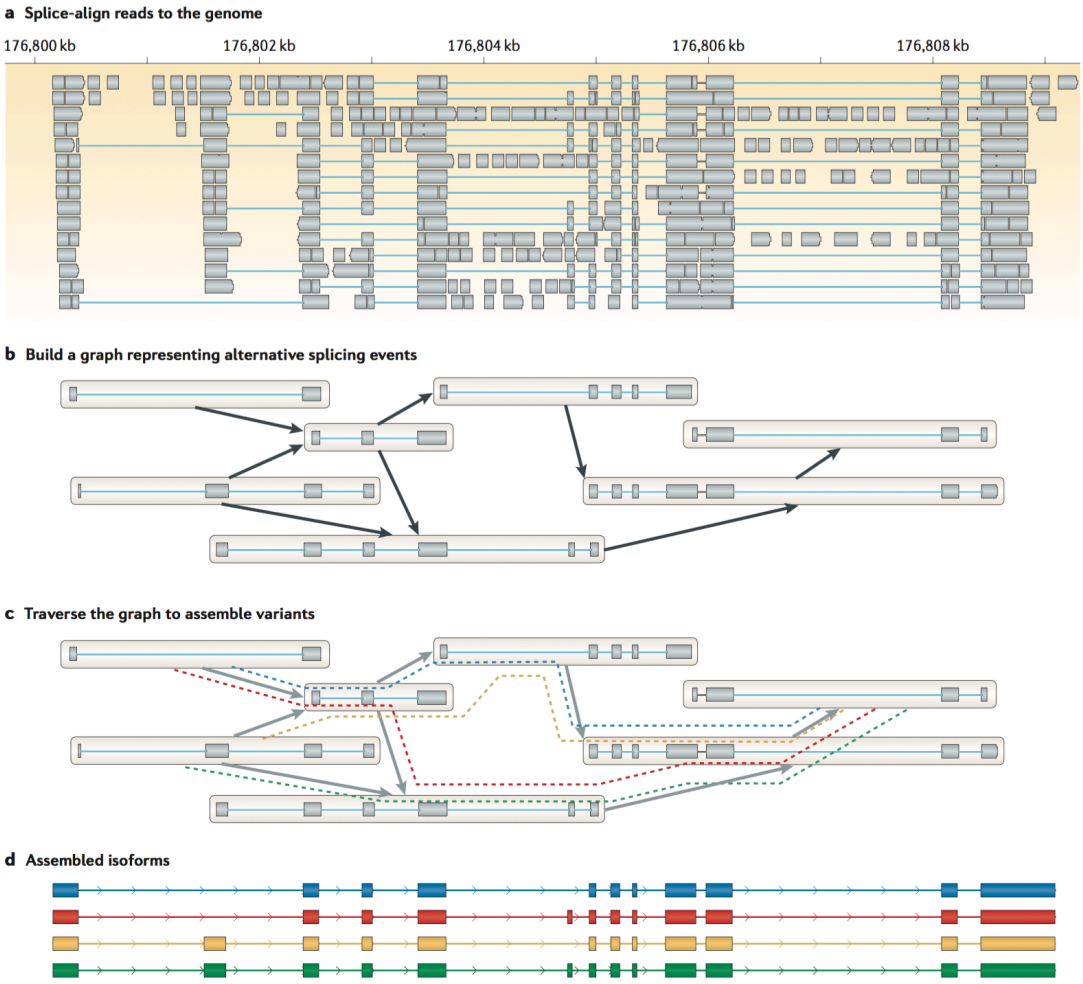


Figure 13: Overview of the reference-based transcriptome assembly strategy – adapted from (Martin and Wang, 2011)

When mapping RNA-seq reads to a genome, the challenge is to identify splice-junctions properly so that junction-spanning reads can be mapped accurately. This is performed by a special type of mappers, called spliced-mappers, which either map reads to the genome as a block using unspliced mapper (allowing no gap opening in the alignment of the read onto the genome) in order to identify exons, and then use the remaining reads to search for splice junctions, called "exon-first" approach used by TopHat (Trapnell et al., 2009), MapSplice (Wang et al., 2010) and SOAPsplice (Huang et al., 2011); or split reads into substrings of a determined size, also called 'seeds', which are mapped in one block without mismatches allowed to the genome, and then extended to join seeds and reconstruct fragments, which is the "seed-and-extend" approach

employed by STAR (Dobin et al., 2013), GSNAP (Wu and Nacu, 2010) and HISAT (Kim et al., 2015) for example (Figure 14).

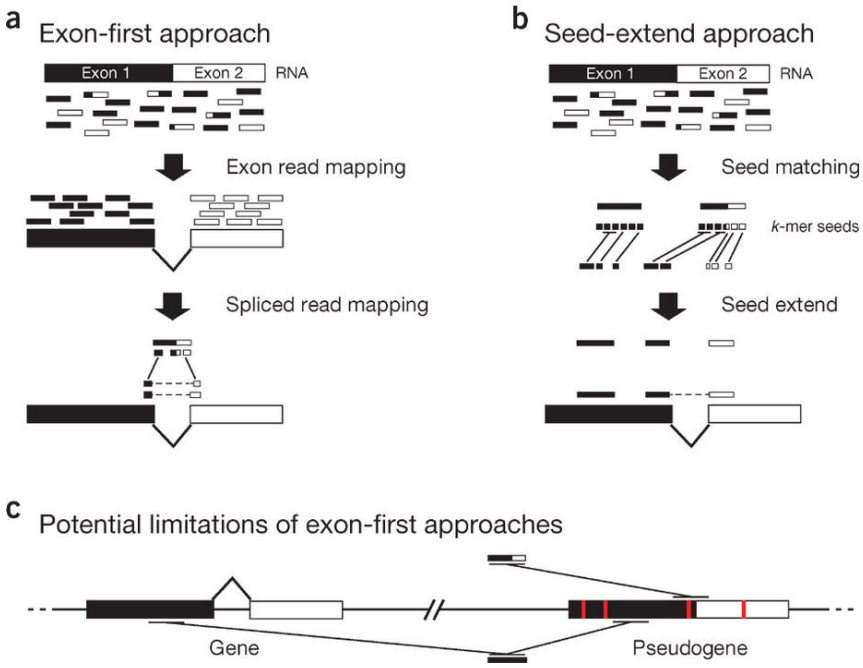


Figure 14: Strategies for gapped alignments of RNA-seq reads to the genome – adapted from (Garber et al., 2011a).

The so-called 'exon-first' approaches are efficient, but less sensitive than the 'seed-and-extend' methods which are more computational intensive, but are better adapted for the identification of novel splice-junctions (Garber et al., 2011b; Engstrom et al., 2013). Among the exon-first and seed-and-extend mapping approaches, some tools like TopHat and STAR can perform transcripts inference with the support of annotation files recording gene models and previously described spliced-junctions, in order to help the discovery of novel splicing events. A third strategy worth mentioning is employed by CRAC (Philippe et al., 2013) and uses k-mer profiling at each genomic location and allows to detect not only splice-junction but also SNP variants and INDELs. Once RNA-seq reads are aligned to the genome, there are two main methods for connecting reads and reconstructing transcripts isoform sequences. The first of them creates a graph of reads overlapping at a single locus (gene), and then traverses the graph to assemble isoforms. Isoforms are assembled by finding the minimum set of transcripts that 'explain' the intron junctions within the reads (that is, a minimum path cover of the graph). This strategy is employed by Cufflinks

(Trapnell et al., 2012). The second method is used by Scripture (Guttman et al., 2010) and first builds a connectivity graph by drawing an edge between any two bases that are connected by a spliced read gap, and then scans the graph with coverage from all reads (spliced and unspliced) to report only path with significant coverage. As a result, Cufflinks is said to be more conservative in its choice of which transcripts to reconstruct, whereas Scripture may produce a larger set of transcripts from a locus (Martin and Wang, 2011).

The reference-based approach has the advantage of allowing the assembly of lowly abundant and novel transcripts that are not present in the available annotation. As the underlying genome sequence is known, small gaps within the transcript sequence due to the lack of read coverage can be filled in using the reference sequence (Roberts et al., 2011). Another advantage of this strategy is that contamination or sequencing artefacts are not a major concern, because they are not expected to align to the reference genome. However, the success of reference-based assemblers depends on the quality of the reference genome. Gaps and miss-assemblies in the reference genome may mislead the assembler and result in misassembled or partially assembled transcriptomes. Although the *ab-initio* strategy is not possible without a reference genome, it is possible in some cases to use the reference genome of a closely related species. Obviously, the accuracy of the transcripts assembly will reflect the degree of synteny between the two species.

2.3.3.1.3 De novo

When a reference genome is either not available, or incomplete, RNA-seq reads can be assembled *de novo* into a transcriptome. As for genome assembly, *de novo* transcriptome assembly requires redundancy of short-read sequences to find overlaps between the reads and assembles them into transcripts. As NGS technologies produce large amounts of data set, transcripts cannot be reconstructed directly from the reads as this would be too computer resource demanding. Instead, substrings of a determined length k (k-mers) are generated for each read, which are then assembled into a De Bruijn graph by shifting along the read sequence with a step of l base pair determining an overlap of k-l between two k-mers. In the De Bruijn graph, each SNP or sequencing error opens a bubble of length k, while introns and deletions introduce a shorter path

in the graph. In a second step, k-mers are collapsed into contigs, where each unique k-mer represents a node, and nodes are connected by overlapping k-mers. Finally, as for genome-guided assemblies, the graph is traversed to reconstruct all alternative paths, and transcripts isoforms are assembled (Figure 15).

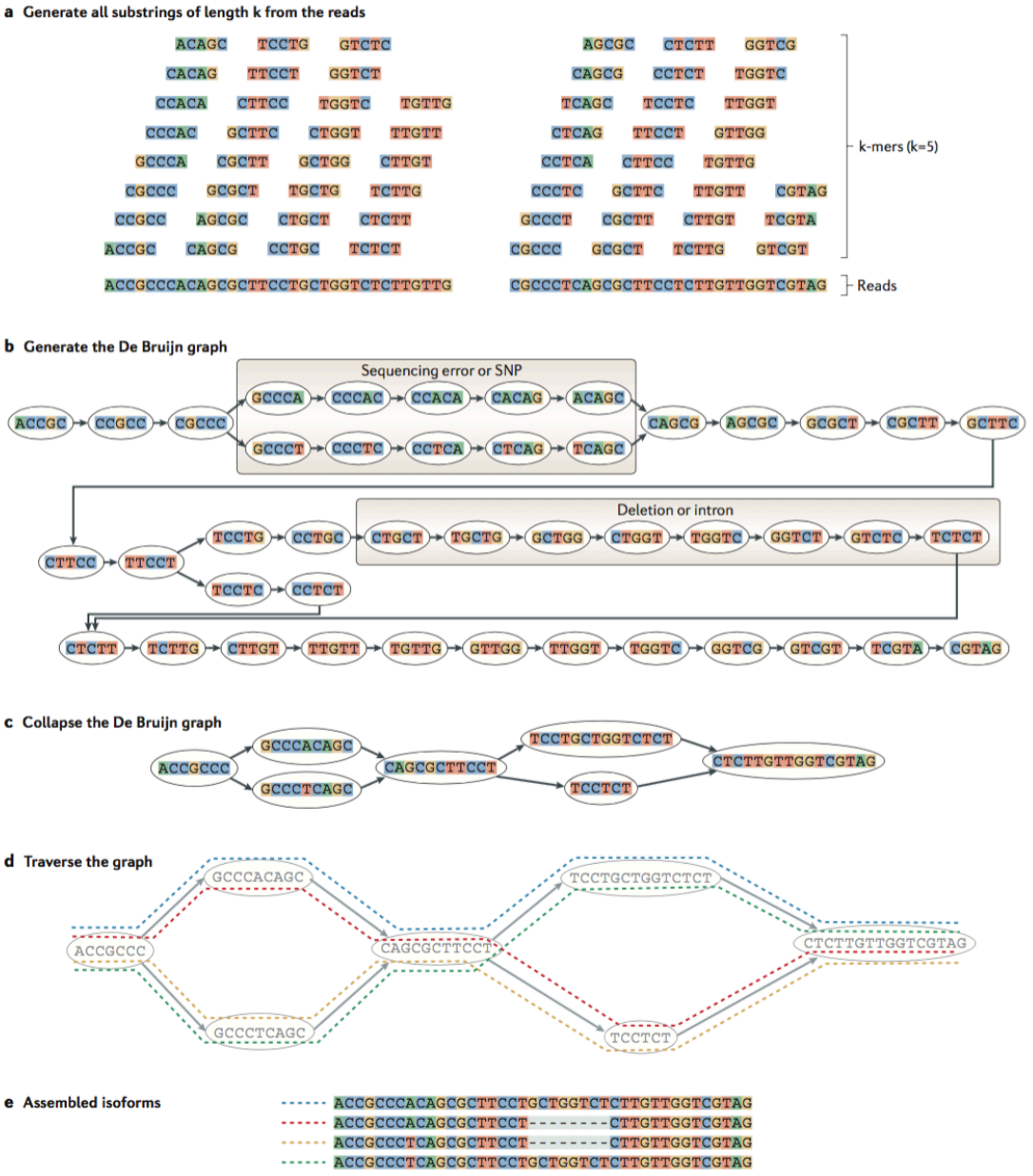


Figure 15: Overview of the de novo *transcriptome assembly strategy- adapted from (Martin and Wang, 2011).*

De novo transcriptome assembly approach has two main advantages. First, it does not depend on a reference genome and thus can be used for any organism. Second it does not rely on accurate

read mapping to identify transcript isoforms and thus enables discovery of novel splice events even in annotated genomes.

A large variety of computer programs has been developed for *de novo* transcriptome assembly (Table 3), including SOAPdenovo-Trans (Xie et al., 2014), Oases (Schulz et al., 2012), Trans-ABySS (Robertson et al., 2010; Grabherr et al., 2011) and Trinity (Grabherr et al., 2011). Most of them are extensions of genome assemblers with which they share the k-mer generation and the first contig assembly steps. However, *de novo* assembly of transcriptome is hampered by several difficulties linked with the intrinsic characteristics of transcripts. Indeed, one RNA-seq read can originate from a large number of transcript isoforms, or genes from the same gene family. The potential multiple origin of a read complicates the reconstruction of transcripts and is the source of chimeras originating from two transcripts sharing common domain(s). In general, paired-end and strand-specific sequencing as well as long reads are preferred in *de novo* transcriptome assembly because they are more informative and allow the reconstruction of full length transcripts (Conesa et al., 2016).

Table 3: Software for De Novo transcriptome assembly – adapted from (Lowe et al., 2017b).

Software (Manufacturer)	Released	Last Updated	Resource load	Strengths and weaknesses
Velvet-Oases [100][101]	2008	2011	Heavy	The original short read assembler, now largely superseded.
SOAPdenovo-trans [102]	2011	2015	Moderate	Early short read assembler, updated for transcript assembly.
Trans-ABySS [103]	2010	2016	Moderate	Short reads, large genomes, MPI-parallel version available.
Trinity [104][105]	2011	2017	Moderate	Short reads, large genomes, memory intensive.
miraEST [106]	1999	2016	Moderate	Repetitive sequences, hybrid data input, wide range of sequence platforms accepted.
Newbler [107]	2004	2012	Heavy	Specialised for Roche 454 sequence, homo-polymer error handling.
CLC genomics workbench (Qiagen—Venlo, Netherlands) [108]	2008	2014	Light	Graphical user interface, hybrid data.

MPI, Message Passing Interface; RNA-Seq, RNA sequencing.

https://doi.org/10.1371/journal.pcbi.1005457.t003

A second limitation is the difficulty to assemble lowly expressed transcripts, as they lack enough coverage for reliable assembly. Transcriptome sequence coverage depth is generally uneven and varies between genes and isoforms according their expression level. Therefore, one nucleotide variant must be supported by enough sequencing depth in order to be resolved from sequencing

error. However, it has also been shown that too many reads are also problematic because they lead to potential miss-assemblies and increased runtimes. Therefore, *in silico* reduction of the number of reads is recommended for deeply sequenced samples (Haas et al., 2013; Conesa et al., 2016).

Different transcriptome reconstruction approaches yield different levels of transcript information. In a 'transcriptome' mapping approach, mapping against the transcriptome only prevents the discovery of new, unannotated transcripts and makes the analysis only quantitative. In the 'reference-based' approach, novel and already described transcripts can be studied, but the quality of the assembly relies on the quality of annotation. Finally, although the '*de novo*' approach overcomes the need for a reference genome or transcriptome, it is important to be aware of the fragmented nature of the transcriptome assembly in which a certain fraction of assembled contigs may not be full-length transcripts. One must therefore take these characteristics into consideration for the interpretation of the results.

2.3.3.2 Statistical analysis of transcriptome and other applications

Once the transcriptome of a given sample is reconstructed, one of the main applications of RNA-seq is the quantification of transcript abundance. Although the genome is the same in all cells of an organism, the composition and abundance of transcripts may differ depending on cell type, tissue, and physiological conditions (developmental stage, stress, etc.). The aim is therefore to quantify each transcript in different samples and then compare the transcript abundance.

2.3.3.2.1 Transcripts abundance estimation

The abundance of a transcript in a sample is directly proportional to the number of reads produced by RNA sequencing. In theory, the most straightforward method to quantify gene expression is to 'count' the number of reads mapping to a genome locus. In reality, however, genes can express more than one transcript and the expression of a gene is the sum of expression of all its transcripts. Furthermore, some reads will map equally well to multiple locations as a result of conserved domains between members of a gene family, between paralogous and/or homeologous genes, or

because several genic features overlap at one locus. This 'read assignment uncertainty' of multi-mapped reads affects expression quantification accuracy.

Two main methods have been developed for quantification of gene expression (Figure 16A): the 'transcript-based' approach, also known as 'exon intersection' approach, such as RSEM (Li and Dewey, 2011), Cufflinks (Trapnell et al., 2012), eXpress (Roberts and Pachter, 2013), isoEM (Nicolae et al., 2011) and the 'union exon'-based approaches such as HTSeq-count (Anders et al., 2015) and featureCounts (Liao et al., 2014). The transcript-based approach relies on expectation-maximization algorithm (Dempster et al., 1977) to assign reads to transcript isoforms. This approach can deal with reads aligning to multiple genomic loci and those aligning to loci for which more than one feature (gene or exon) is annotated. Methods such as RSEM or eXpress solve the read assignment uncertainty issue by normalizing the counts of multi-mapped reads according the relative abundances of the transcripts estimated from uniquely mapped reads (Figure 16B).

However, the isoform quantification may encounter difficulties if important proportion of transcripts overlaps. Informative reads such as paired-end reads are thus preferred for these algorithms to help resolving the transcript origin of multi-mapped reads.

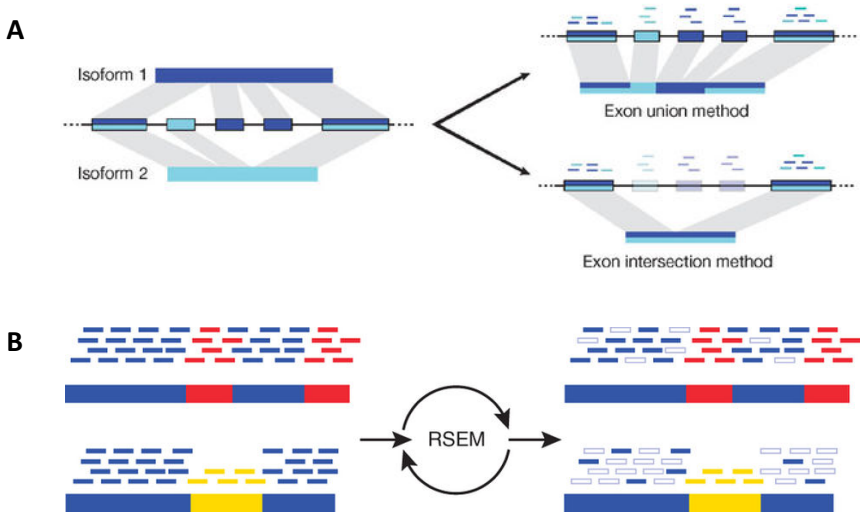


Figure 16: Methods of gene expression quantification: (A) The exon 'union' and 'intersection' methods – adapted from (Garber et al., 2011a); (B) The RNA-Seq by Expectation Maximization (RSEM) method – adapted from (Haas et al., 2013).

In contrast, the ‘union exon’-based counting method considers genes as unions of their exons and thus merges all overlapping exons of the same gene into ‘union exons’, and then counts reads for each gene by intersecting all mapped reads with ‘union exons’ of the gene. Calculating genes or exon read counts is generally easier as reads can be assigned with higher confidence. Therefore, the ‘union exons’-based counting method is widely used, even if gene level counts with this method do not distinguish between isoforms when multiple transcripts are expressed from the same gene. Finally, the ‘union exons’ method has limitations when dealing with multi-mapped reads, as it only allows to discard those reads or count them multiple times. To conclude, the “union based’ methods seem to overestimates gene expression (as reads may be counted multiple times), while ‘transcripts based’ methods tend to underestimate the true change in gene expression (Trapnell et al., 2012).

Some quantification software circumvent the need for exact alignment of reads to a reference sequence. Instead, they determine the compatibility of reads with targets based on k-mers alignment like Sailfish (Patro et al., 2014) and RNA-Skim (Zhang and Wang, 2014), or on k-mer ‘pseudoalignments’ like Kallisto (Bray et al., 2016). In k-mer alignment, read sequences are hashed into smaller sequences of *k* size called k-mers that are then mapped back to the transcripts in order to quantify them. The pseudoalignment of k-mers solves the problem the other way around by generating a k-mer De Bruijn graph from the transcriptome, trying to assign a *k-compatibility class* to each k-mer (i.e., associate a k-mer to one or more transcripts), and then attempts to identify transcripts from which the k-mer (and therefore the read) could have originated. These k-mer based quantification methods are faster than the read mapping-based methods, produce similar results as read mapping-based methods and require less computational resources (Borrill et al., 2016).

2.3.3.2.2 Transcripts expression normalization

Although raw reads counts are useful for differential expression analysis, they have little relevance as ‘expression’ unit. Indeed, when using RNA-seq to quantify transcripts expression, two main sources of variability are systematically introduced in the final raw read count. The first

source of variation comes from RNA or cDNA fragmentation during the library preparation, which causes longer transcripts to generate more reads compared to shorter transcripts present at the same abundance level (Figure 17). Second, different sequencing runs will produce different number of reads, which causes the same transcript to produce different number of reads as a result of sequencing depth and not differential expression. In order to account for these sources of variability and obtain meaningful expression estimates, several metrics have been developed that normalize raw counts.

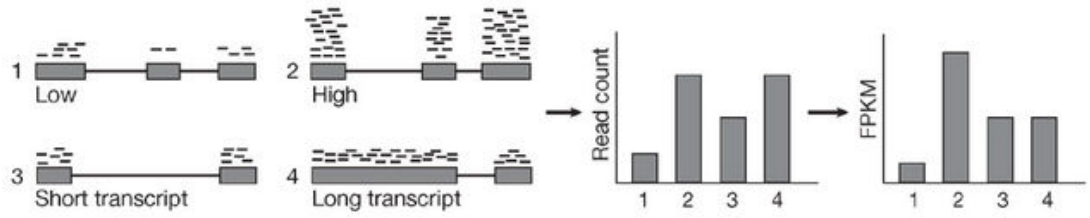


Figure 17: Impact of sequencing depth and transcripts length on raw (Read count) and normalized (FPKM) expression values – adapted from (Garber et al., 2011b).

'Counts per million' (CPM) mapped reads are raw read counts normalized by the total number of sequencing reads produced from a sample, called library size, and allow for comparison of the same gene or transcript between two or more samples. The 'Reads/Fragments Per Kilobase of transcripts per Million mapped reads' (RPKM/FPKM) metrics are transcript expression values normalized for the library size in case of single-end or paired-end sequencing, respectively, as well as by the size of the transcripts (Mortazavi et al., 2008). These two metrics permit comparison of the expression of the same gene between samples and/or the comparison of the expression of different genes within the same sample, and are the most widely used expression units in RNA-seq analysis (Garber et al., 2011b). Finally, the third metrics, 'Transcript Per Million mapped reads' (TPM) normalizes for differences in the composition of transcripts rather than simply dividing by the number of reads in the library (Li and Dewey, 2011). TPM values are considered more suitable for comparisons between samples of different origin and composition (Conesa et al., 2016).

2.3.3.2.3 Differential expression analysis

The aim of Differential Expression (DE) analysis is to test whether the level of gene expression is significantly different between two samples. Once quantitative counts of each transcript are available, differential gene expression is measured by normalizing, modelling, and statistically analysing the data set. Various computer programs have been developed for the differential expression analysis of RNA-seq data. However, DESeq (Anders and Huber, 2010) and edgeR (Robinson et al., 2010) are currently the most popular. Before comparing two samples, it is necessary to normalize the counts to make them comparable. DE analysis software normalizes the data for principally two sample-specific sources of variation affecting the counts: the sequencing depth, which is demonstrated by differences in library size and RNA composition. Difference in RNA composition may occur for example when a small number of genes are highly expressed in one sample, but not in another. The highly expressed genes can consume a substantial proportion of the total sequencing library size, causing the remaining genes to be under-sampled. Unless this RNA composition effect is adjusted for, the remaining genes may falsely appear to be down-regulated in that sample.

While the effect of sequencing depth is normalized by dividing raw counts by the library size or the number of transcripts, variation in RNA composition between samples is typically normalized by two methods: the Trimmed mean of M-values (TMM) method provided by edgeR software (Robinson and Oshlack, 2010) and the "Relative Log Expression" normalization (RLE) method from DESeq (Anders and Huber, 2010). Both methods first estimate the "relative RNA production of two samples" which corresponds to the ratio of overall gene expression between two samples, and allow for the estimation of RNA "library composition" divergence between these two samples, namely the 'scale factor'. In order to ensure that a gene with the same expression level in two samples is not detected as DE, this scale factor is applied to normalize the library sizes, and thus make counts from the two samples comparable. Both methods assume that most genes are not differentially expressed. As GC-content and gene/transcript length are not different when comparing two samples, these parameters do not need to be normalized in DE analysis.

Due to low number of biological replicates in RNA-seq analyses, differential expression cannot be directly inferred from count-based distribution such as the Poisson distribution as these distributions do not account for biological variability across samples (Garber et al., 2011b). To overcome this, edgeR and DESeq differential expression methods use a negative binomial distribution to model biological variability and provide a measure of significance in the absence of a large number of replicates. Once counts are modelled, the DE is tested for each gene using the Fisher exact test (in case of monoparametric experiments where only one parameter is expected to affect gene expression, e.g., stress or time) or the likelihood ratio test (in case of multi-parametric experiments where more than one parameter is expected to affect gene expression, e.g., stress and time). A *p*-value is assigned to each gene tested and corresponds to the probability that the difference observed between conditions is not more extreme than the difference expected under the null hypothesis of identical distribution counts in the two conditions.

2.3.3.3 Functional profiling

Identification of sets of differentially expressed genes or transcripts is often not the final goal of the study, but a milestone in the effort to understand a biological mechanism. The last step in a standard transcriptomics study is thus characterization of molecular functions or biological pathways in which differentially expressed genes (DGE) are involved.

One of the first approaches to characterize biological importance of DGE is to investigate their functional annotation. Databases such as Gene Ontology (Ashburner et al., 2000), DAVID (Dennis et al., 2003; Al-Shahrour et al., 2005) and Babelomics (Al-Shahrour et al., 2005) contain gene annotation data for a majority of model species. However, as RNA-seq analysis is also meant to discover novel transcripts, it is possible that for certain transcripts, no functional information is available. In these cases, functional annotation can be performed by orthology search for similar sequences. For protein coding genes, the search can be performed against protein and protein-domains databases such as SwissProt (Bairoch and Apweiler, 2000), UniProt

(Apweiler et al., 2004), Pfam (Apweiler et al., 2001; Bateman et al., 2002) and InterPro (Apweiler et al., 2001). For non-coding RNA, specialized databases can be accessed such as Rfam (Griffiths-Jones et al., 2003), mirBase (Griffiths-Jones, 2006) or Miranda (Enright et al., 2003). Accessing the functional annotation of DGEs is an important step towards the understanding of their biological relevance as it allows to search literature for their functions. However, it involves an overwhelming amount of information as differential expression analysis often yields several hundreds of DGEs. As it is not always possible to investigate each gene individually, an approach widely used to characterise the biological functions of a set of DGEs is to perform a Gene Ontology (GO) enrichment analysis (http://geneontology.org/). This type of analysis makes the use of gene annotation to find over- or under-representation of GO terms in the DGEs set, which are functional terms describing molecular function, biological process or cellular compartment associated with each gene. Several tools have been developed for this analysis such as Blast2GO (Conesa et al., 2005), BiNGO (Maere et al., 2005) and Ontologizer (Bauer et al., 2008). Other applications of RNA-seq include the study of alternative splicing, gene fusion, or variant analysis.

2.3.4 Transcriptome analysis of products of inter-specific (-generic) hybridization

To date, several transcriptomic studies have been performed in products of interspecific-hybridization in rice (Wu et al., 2016; Guo et al., 2017), brassicas (Chu et al., 2014; Zhang et al., 2015), ryegrass (Stoces et al., 2016) and newly formed allopolyploids such as synthetic wheats (Wang et al., 2016). These experiments generally aimed at assessing parental genome dominance and heterosis effects. In contrast, there is a large gap in understanding alien gene regulation in alien introgression lines and how the alien genes affect host genome functions. Cho et al. (2006) used six out of the seven wheat Chinese Spring - barley Betzes disomic addition lines (Islam et al., 1981) and thirteen of the fourteen ditelosomic addition lines (Islam, 1983) produced by Islam and colleagues containing a complete chromosome set (42 chromosome) of hexaploid wheat and a single pair of barley chromosomes (-arms) to perform the first transcriptome analysis using the Barley Affymetrix GeneChip probe array. The authors were able to identify chromosome origin of 1,787 barley transcripts from chromosomes 1HS, and 2HS - 7HL. However, this represented

only a limited fraction of the full complement of transcripts, estimated to 39,734 and did not allow to estimate gene expression levels in each genotype (Mascher et al., 2017).

Clearly, RNA-seq seems more appropriate for such analysis. However, its application in wheat-barley introgression material is hampered by difficulties related to multiple levels of gene homology. Polyploid crop species such as wheat contain closely related homeologous sub-genomes, which share highly similar nucleotide sequences within coding regions. This poses a challenge for RNA-seq analysis in which assigning short reads to the correct gene copy (homeolog) implies a certain degree of uncertainty. Bread wheat contains three copies of a majority of genes, one on each of the A, B and D homeologous chromosomes, and these genes share over 95 % identity in coding sequences (Krasileva et al., 2013). Apart from the high level of sequence identity between gene homeologues, wheat and barley share extensive conservation of genome synteny (Moore et al., 1995; Mayer et al., 2011a). Nevertheless, recent studies showed that the rate of polymorphism between homologous sequences combined with stringent read mapping and transcripts quantification strategies allow for RNA-seq to accurately discriminate between homeologous genes expression with only 5-6% of residual expression (Krasileva et al., 2013; Leach et al., 2014; Borrill et al., 2016). These results are far better than those obtained using hybridization-based methods such as microarrays and open avenues for comprehensive transcriptome analysis in wheat-alien introgression lines.

References

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F. and et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651-1656.

Al-Shahrour, F., Minguez, P., Vaquerizas, J.M., Conde, L. and Dopazo, J. (2005) BABELOMICS: a suite of web tools for functional annotation and analysis of groups of genes in high-throughput experiments. *Nucleic Acids Res* **33**, W460-464.

Alberts B, J.A., Lewis J, et al. (2002) *Molecular Biology of the Cell. 4th edition.* New York:
Garland Science.

Alexandratos, N. and Bruinsma, J. (2012) World agriculture towards 2030/2050: the 2012 revision. ESA Working paper FAO, Rome.

Allendorf, D.J., Yan, J., Ross, G.D., Hansen, R.D., Baran, J.T., Subbarao, K., Wang, L. and Haribabu, B. (2005) C5a-mediated leukotriene B4-amplified neutrophil chemotaxis is essential in tumor immunotherapy facilitated by anti-tumor monoclonal antibody and β-glucan. *The Journal of Immunology* **174**, 7050-7056.

Alwine, J.C., Kemp, D.J. and Stark, G.R. (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *P Natl Acad Sci USA* **74**, 5350-5354.

Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome biology* **11**, R106.

Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)* **31**, 166-169.

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Hermjakob, H., Hulo, N., Jonassen, I., Kahn, D., Kanapin, A., Karavidopoulou, Y., Lopez, R., Marx, B., Mulder, N.J., Oinn, T.M., Pagni, M., Servant, F., Sigrist, C.J. and Zdobnov, E.M. (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res* **29**, 37-40.

Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N. and Yeh, L.S. (2004) UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* **32**, D115-119.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M. and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29.

Autrique, E., Tanksley, S.D., Sorrells, M.E. and Singh, R.P. (1995) Molecular markers for four leaf rust resistance genes introgressed into wheat from wild relatives. *Genome* **38**, 75-83.

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

Bairoch, A. and Apweiler, R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* **28**, 45-48.

Baruch, K., Barad, O., Ben Zvi, G. and Ronen, G. (2016) *Next generation complex genome assembly*.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and Sonnhammer, E.L. (2002) The Pfam protein families database. *Nucleic Acids Res* **30**, 276-280.

Bauer, S., Grossmann, S., Vingron, M. and Robinson, P.N. (2008) Ontologizer 2.0--a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics* **24**, 1650-1651.

Bennetzen, J.L. (2000) Comparative Sequence Analysis of Plant Nuclear Genomes: Microcolinearity and Its Many Exceptions. *The Plant cell* **12**, 1021-1029.

Bennetzen, J.L. (2002) Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* **115**, 29-36.

Bilgic, H., Cho, S., Garvin, D.F. and Muehlbauer, G.J. (2007) Mapping barley genes to chromosome arms by transcript profiling of wheat–barley ditelosomic chromosome addition lines. *Genome* **50**, 898-906.

Blakeslee, A.F. (1937) Dédoublement du nombre de chromosomes chez les plantes par traitement chimique. *Comptes rendus des séances de l'Académie des Sciences* **205**, 476.

Bolot, S., Abrouk, M., Masood-Quraishi, U., Stein, N., Messing, J., Feuillet, C. and Salse, J. (2009) The 'inner circle'of the cereal genomes. *Current opinion in plant biology* **12**, 119-125.

Börner, A., Ogbonnaya, F.C., Röder, M.S., Rasheed, A., Periyannan, S. and Lagudah, E.S. (2015) Aegilops tauschii introgressions in wheat. In: *Alien Introgression in Wheat* pp. 245-271. Springer.

Borrill, P., Ramirez-Gonzalez, R. and Uauy, C. (2016) expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform. *Plant Physiology* **170**, 2172-2186.

Borst, P. (1986) Discontinuous Transcription and Antigenic Variation in Trypanosomes. *Annu Rev Biochem* **55**, 701-732.

Bray, N.L., Pimentel, H., Melsted, P. and Pachter, L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-527.

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G.L., D'Amore, R., Allen, A.M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M., Bancroft, I., Gu, Y., Huo, N., Luo, M.C., Sehgal, S., Gill, B., Kianian, S., Anderson, O., Kersey, P., Dvorak, J., McCombie, W.R., Hall, A., Mayer, K.F., Edwards, K.J., Bevan, M.W. and Hall, N. (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491**, 705-710.

Brunner, S., Hurni, S., Herren, G., Kalinina, O., von Burg, S., Zeller, S.L., Schmid, B., Winzeler, M. and Keller, B. (2011) Transgenic Pm3b wheat lines show resistance to powdery mildew in the field. *Plant biotechnology journal* **9**, 897-910.

Bustin, S.A. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of molecular endocrinology* **25**, 169-193.

Cao, J. (2014) The functional role of long non-coding RNAs and epigenetics. *Biol Proced Online* **16**, 11.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V.B., Brenner, S.E., Batalov, S., Forrest, A.R., Zavolan, M., Davis, M.J., Wilming, L.G., Aidinis, V., Allen, J.E., Ambesi-Impiombato, A., Apweiler, R., Aturaliya, R.N., Bailey, T.L., Bansal, M., Baxter, L., Beisel, K.W., Bersano, T., Bono, H., Chalk, A.M., Chiu, K.P., Choudhary, V., Christoffels, A., Clutterbuck, D.R., Crowe, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C.F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T.R., Gojobori, T., Green, R.E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T.K., Hirokawa, N., Hill, D., Huminiecki, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S.P., Kruger, A., Kummerfeld, S.K., Kurochkin, I.V., Lareau, L.F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Morris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N.,

Nakauchi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, K.C., Pavan, W.J., Pavesi, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J.F., Ring, B.Z., Ringwald, M., Rost, B., Ruan, Y., Salzberg, S.L., Sandelin, A., Schneider, C., Schonbach, C., Sekiguchi, K., Semple, C.A., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, E., Sugiura, K., Sultana, R., Takenaka, Y., Taki, K., Tammoja, K., Tan, S.L., Tang, S., Taylor, M.S., Tegner, J., Teichmann, S.A., Ueda, H.R., van Nimwegen, E., Verardo, R., Wei, C.L., Yagi, K., Yamanishi, H., Zabarovsky, E., Zhu, S., Zimmer, A., Hide, W., Bult, C., Grimmond, S.M., Teasdale, R.D., Liu, E.T., Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J.S., Hume, D.A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Konno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M., Plessy, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watahiki, A., Okamura-Oho, Y., Suzuki, H., Kawai, J., Hayashizaki, Y., Consortium, F., Group, R.G.E.R. and Genome Science, G. (2005) The transcriptional landscape of the mammalian genome. *Science* **309**, 1559-1563.

Castillo, A., Atienza, S.G. and Martín, A.C. (2014) Fertility of CMS wheat is restored by two Rf loci located on a recombined acrocentric chromosome. *Journal of Experimental Botany* **65**, 6667-6677.

Chapman, J.A., Mascher, M., Buluc, A., Barry, K., Georganas, E., Session, A., Strnadova, V., Jenkins, J., Sehgal, S., Oliker, L., Schmutz, J., Yelick, K.A., Scholz, U., Waugh, R., Poland, J.A., Muehlbauer, G.J., Stein, N. and Rokhsar, D.S. (2015) A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome biology* **16**, 26.

Chen, G., Zheng, Q., Bao, Y., Liu, S., Wang, H. and Li, X. (2012) Molecular cytogenetic identification of a novel dwarf wheat line with introgressed Thinopyrum ponticum chromatin. *Journal of biosciences* **37**, 149-155.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., Sementchenko, V., Piccolboni, A., Bekiranov, S., Bailey, D.K., Ganesh, M., Ghosh, S., Bell, I., Gerhard, D.S. and Gingeras, T.R. (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**, 1149-1154.

Cho, S., Garvin, D.F. and Muehlbauer, G.J. (2006) Transcriptome analysis and physical mapping of barley genes in wheat-barley chromosome addition lines. *Genetics* **172**, 1277-1285.

Choulet, F., Alberti, A., Theil, S., Glover, N., Barbe, V., Daron, J., Pingault, L., Sourdille, P., Couloux, A., Paux, E., Leroy, P., Mangenot, S., Guilhot, N., Le Gouis, J., Balfourier, F., Alaux, M., Jamilloux, V., Poulain, J., Durand, C., Bellec, A., Gaspin, C., Safar, J., Dolezel, J., Rogers, J., Vandepoele, K., Aury, J.M., Mayer, K., Berges, H., Quesneville, H., Wincker, P. and Feuillet, C. (2014) Structural and functional partitioning of bread wheat chromosome 3B. *Science* **345**, 1249721.

Chu, P., Liu, H.J., Yang, Q., Wang, Y.K., Yan, G.X. and Guan, R.Z. (2014) An RNA-seq transcriptome analysis of floral buds of an interspecific Brassica hybrid between B-carinata and B-napus. *Plant Reprod* **27**, 225-237.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics (Oxford, England)* **21**, 3674-3676.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X. and Mortazavi, A. (2016) A survey of best practices for RNA-seq data analysis. *Genome biology* **17**, 13.

Cox, J.M. (2001) Applications of nylon membrane arrays to gene expression analysis. *Journal of Immunological Methods* **250**, 3-13.

Crick, F.H. (1958) On protein synthesis. *Symp Soc Exp Biol* **12**, 138-163.

Cseh, A., Kruppa, K., Molnar, I., Rakszegi, M., Dolezel, J. and Molnar-Lang, M. (2011) Characterization of a new 4BS.7HL wheat-barley translocation line using GISH, FISH,

and SSR markers and its effect on the beta-glucan content of wheat. *Genome* **54**, 795-804.
Cseh, A., Soos, V., Rakszegi, M., Turkoesi, E., Balazs, E. and Molnar-Lang, M. (2013) Expression of HvCslF9 and HvCslF6 barley genes in the genetic background of wheat and their influence on the wheat beta-glucan content. *Annals of Applied Biology* **163**, 142-150.
Del Blanco, I., Rajaram, S. and Kronstad, W. (2001) Agronomic potential of synthetic hexaploid wheat-derived populations. *Crop Science* **41**, 670-676.
Delibes, A., Romero, D., Aguaded, S., Duce, A., Mena, M., Lopez-Braña, I., Andrés, M.-F., Martin-Sanchez, J.-A. and García-Olmedo, F. (1993) Resistance to the cereal cyst nematode (Heterodera avenae Woll.) transferred from the wild grassAegilops ventricosa to hexaploid wheat by a "stepping-stone" procedure. *Theoretical and Applied Genetics* **87**, 402-408.
Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1-38.
Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. and Lempicki, R.A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome biology* **4**, P3.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, 15-21.
Doležel, J., Vrána, J., Šafář, J., Bartoš, J., Kubaláková, M. and Šimková, H. (2012) Chromosomes in the flow to simplify genome analysis. *Functional & Integrative Genomics* **12**, 397-416.
Dubcovsky, J. and Dvorak, J. (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* **316**, 1862-1866.
Egawa, C., Kobayashi, F., Ishibashi, M., Nakamura, T., Nakamura, C. and Takumi, S. (2006) Differential regulation of transcript accumulation and alternative splicing of a DREB2 homolog under abiotic stress conditions in common wheat. *Genes & genetic systems* **81**, 77-91.
Eisen, H. (1988) RNA editing: who's on first? *Cell* **53**, 331-332.
Endo, T. (2007) The gametocidal chromosome as a tool for chromosome manipulation in wheat. *Chromosome Research* **15**, 67-75.
Engstrom, P.G., Steijger, T., Sipos, B., Grant, G.R., Kahles, A., Ratsch, G., Goldman, N., Hubbard, T.J., Harrow, J., Guigo, R., Bertone, P. and Consortium, R. (2013) Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat Methods* **10**, 1185-1191.
Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C. and Marks, D.S. (2003) MicroRNA targets in Drosophila. *Genome biology* **5**, R1.
Eversole, K., Feuillet, C., Mayer, K.F. and Rogers, J. (2014a) Slicing the wheat genome. American Association for the Advancement of Science.
Eversole, K., Feuillet, C., Mayer, K.F. and Rogers, J. (2014b) Slicing the wheat genome. Introduction. *Science* **345**, 285-287.
Farkas, A., Molnár, I., Kiss, T., Karsai, I. and Molnár-Láng, M. (2014) Effect of added barley chromosomes on the flowering time of new wheat/winter barley addition lines in various environments. *Euphytica* **195**, 45-55.
Farrer, W. (1904) Some notes on the wheat "Bobs"; its peculiarities, economic value, and origin. *Agric Gaz NSW* **15**, 849-854.
Fatih, A. (1983) Analysis of the breeding potential of wheat-Agropyron and wheat-Elymus derivatives. *Hereditas* **98**, 287-295.
Fedak, G. (1999) Molecular aids for integration of alien chromatin through wide crosses. *Genome* **42**, 584-591.
Feldman, M. and Levy, A.A. (2015) Origin and Evolution of Wheat and Related Triticeae Species. In: *Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and*

Genomics (Molnár-Láng, M., Ceoloni, C. and Doležel, J. eds), pp. 21-76. Cham: Springer International Publishing.

Feuillet, C., Langridge, P. and Waugh, R. (2008) Cereal breeding takes a walk on the wild side. *Trends in genetics: TIG* **24**, 24-32.

Filichkin, S.A., Cumbie, J.S., Dharmawardhana, P., Jaiswal, P., Chang, J.H., Palusa, S.G., Reddy, A.S., Megraw, M. and Mockler, T.C. (2015) Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in Arabidopsis. *Mol Plant* **8**, 207-227.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811.

Friebe, B., Jiang, J., Raupp, W., McIntosh, R. and Gill, B. (1996) Characterization of wheat-alien translocations conferring resistance to diseases and pests: current status. *Euphytica* **91**, 59-87.

Garber, M., Grabherr, M.G., Guttman, M. and Trapnell, C. (2011a) Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature methods* **8**, 469.

Garber, M., Grabherr, M.G., Guttman, M. and Trapnell, C. (2011b) Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods* **8**, 469-477.

Gerstein, M.B., Bruce, C., Rozowsky, J.S., Zheng, D.Y., Du, J., Korbel, J.O., Emanuelsson, O., Zhang, Z.D.D., Weissman, S. and Snyder, M. (2007) What is a gene, post-ENCODE? History and updated definition. *Genome Res* **17**, 669-681.

Gill, B.S., Friebe, B.R. and White, F.F. (2011) Alien introgressions represent a rich source of genes for crop improvement. *Proc Natl Acad Sci U S A* **108**, 7657-7658.

Gottlieb, A., Muller, H.G., Massa, A.N., Wanjugi, H., Deal, K.R., You, F.M., Xu, X., Gu, Y.Q., Luo, M.C., Anderson, O.D., Chan, A.P., Rabinowicz, P., Devos, K.M. and Dvorak, J. (2013) Insular organization of gene space in grass genomes. *PloS one* **8**, e54101.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644-652.

Griffiths, S., Sharp, R., Foote, T.N., Bertin, I., Wanous, M., Reader, S., Colas, I. and Moore, G. (2006) Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. *Nature* **439**, 749-752.

Griffiths-Jones, S. (2006) miRBase: the microRNA sequence database. *Methods in molecular biology (Clifton, N.J.)* **342**, 129-138.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. and Eddy, S.R. (2003) Rfam: an RNA family database. *Nucleic Acids Res* **31**, 439-441.

Guo, H., Mendrikahy, J.N., Xie, L., Deng, J., Lu, Z., Wu, J., Li, X., Shahid, M.Q. and Liu, X. (2017) Transcriptome analysis of neo-tetraploid rice reveals specific differential gene expressions associated with fertility and heterosis. *Sci Rep* **7**, 40139.

Gupta, P., Varshney, R.K., Sharma, P. and Ramesh, B. (1999) Molecular markers and their applications in wheat breeding. *Plant breeding* **118**, 369-390.

Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C., Rinn, J.L., Lander, E.S. and Regev, A. (2010) Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol* **28**, 503-510.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N. and Regev, A. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494-1512.

Hajjar, R. and Hodgkin, T. (2007) The use of wild relatives in crop improvement: A survey of developments over the last 20 years. *Euphytica* **156**, 1-13.

Hamilton, A.J. and Baulcombe, D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950-952.
Heller, M.J. (2002) DNA microarray technology: devices, systems, and applications. *Annu Rev Biomed Eng* **4**, 129-153.
Heslop-Harrison, J. and Schwarzacher, T. (2011) Organisation of the plant genome in chromosomes. *The Plant Journal* **66**, 18-33.
Hrdlickova, R., Toloue, M. and Tian, B. (2017) RNA-Seq methods for transcriptome analysis. *Wiley Interdisciplinary Reviews: RNA* **8**.
Huang, S., Zhang, J., Li, R., Zhang, W., He, Z., Lam, T.W., Peng, Z. and Yiu, S.M. (2011) SOAPsplice: Genome-Wide ab initio Detection of Splice Junctions from RNA-Seq Data. *Frontiers in genetics* **2**, 46.
Hurni, S., Brunner, S., Stirnweis, D., Herren, G., Peditto, D., McIntosh, R.A. and Keller, B. (2014) The powdery mildew resistance gene Pm8 derived from rye is suppressed by its wheat ortholog Pm3. *The Plant journal : for cell and molecular biology* **79**, 904-913.
International Barley Genome Sequencing, C., Mayer, K.F., Waugh, R., Brown, J.W., Schulman, A., Langridge, P., Platzer, M., Fincher, G.B., Muehlbauer, G.J., Sato, K., Close, T.J., Wise, R.P. and Stein, N. (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**, 711-716.
International Brachypodium, I. (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* **463**, 763-768.
International Rice Genome Sequencing, P. (2005a) The map-based sequence of the rice genome. *Nature* **436**, 793.
International Rice Genome Sequencing, P. (2005b) The map-based sequence of the rice genome. *Nature* **436**, 793-800.
International Wheat Genome Sequencing, C. (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science* **345**, 1251788.
Islam, A. and Shepherd, K. (1990) Incorporation of barley chromosomes into wheat. In: *Wheat* pp. 128-151. Springer.
Islam, A., Shepherd, K. and Sparrow, D. (1976) Addition of individual barley chromosomes to wheat. In: *Proceedings of the International Barley Genetics Symposium*.
IWGSC (2018) Shifting the limits in wheat research and breeding through a fully annotated and anchored reference genome sequence. *Science*.
Janick, J. (2008) *Plant breeding reviews*:John Wiley & Sons.
Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S. and Gingeras, T.R. (2005) Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res* **15**, 987-997.
Kato, K., Kidou, S.-i. and Miura, H. (2008) Molecular cloning and mapping of casein kinase 2 alpha and beta subunit genes in barley. *Genome* **51**, 208-215.
Kerber, E. (1991) Stem-rust resistance in'Canthatch'hexaploid wheat induced by a nonsuppressor mutation on chromosome 7DL. *Genome* **34**, 935-939.
Kerber, E. and Dyck, P. (1973) Inheritance of stem rust resistance transferred from diploid wheat (Triticum monococcum) to tetraploid and hexaploid wheat and chromosome location of the gene involved. *Canadian Journal of Genetics and Cytology* **15**, 397-409.
Kerckhoffs, D.A., Hornstra, G. and Mensink, R.P. (2003) Cholesterol-lowering effect of β-glucan from oat bran in mildly hypercholesterolemic subjects may decrease when β-glucan is incorporated into bread and cookies. *The American journal of clinical nutrition* **78**, 221-227.
Kihara, H. (1937) Genomanalyse bei Triticum und Aegilops VII. Kurze ubersicht uber die Ergebnisse der Jahre 1934-36. *Mem. Coll. Agric. Kyoto Uuiv.* **41**, 1-61.
Kilian, B., Martin, W. and Salamini, F. (2010) Genetic Diversity, Evolution and Domestication of Wheat and Barley in the Fertile Crescent. In: *Evolution in Action: Case studies in Adaptive Radiation, Speciation and the Origin of Biodiversity* (Glaubrecht, M. ed) pp. 137-166. Berlin, Heidelberg: Springer Berlin Heidelberg.
Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360.

Kleinhofs, A., Kilian, A., Maroof, M.S., Biyashev, R.M., Hayes, P., Chen, F., Lapitan, N., Fenwick, A., Blake, T. and Kanazin, V. (1993) A molecular, isozyme and morphological map of the barley (Hordeum vulgare) genome. *Theoretical and Applied Genetics* **86**, 705-712.

Kornberg, R.D. (2007) The molecular basis of eukaryotic transcription. *Proceedings of the National Academy of Sciences* **104**, 12955-12961.

Krasileva, K.V., Buffalo, V., Bailey, P., Pearce, S., Ayling, S., Tabbita, F., Soria, M., Wang, S., Consortium, I., Akhunov, E., Uauy, C. and Dubcovsky, J. (2013) Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome biology* **14**, R66.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147-157.

Kruse, A. (1973) Hordeum × Triticum hybrids. *Hereditas* **73**, 157-161.

Kung, J.T.Y., Colognori, D. and Lee, J.T. (2013) Long Noncoding RNAs: Past, Present, and Future. *Genetics* **193**, 651-669.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25.

Leach, L.J., Belfield, E.J., Jiang, C.F., Brown, C., Mithani, A. and Harberd, N.P. (2014) Patterns of homoeologous gene expression shown by RNA sequencing in hexaploid bread wheat. *Bmc Genomics* **15**.

Lee, J.H., Ryu, H.S., Chung, K.S., Pose, D., Kim, S., Schmid, M. and Ahn, J.H. (2013) Regulation of temperature-responsive flowering by MADS-box transcription factor repressors. *Science* **342**, 628-632.

Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-854.

Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **12**, 323.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics (Oxford, England)* **25**, 1754-1760.

Li, J., WAN, H.S. and YANG, W.Y. (2014) Synthetic hexaploid wheat enhances variation and adaptive evolution of bread wheat in breeding processes. *Journal of Systematics and Evolution* **52**, 735-742.

Li, L.F., Liu, B., Olsen, K.M. and Wendel, J.F. (2015) A re-evaluation of the homoploid hybrid origin of Aegilops tauschii, the donor of the wheat D-subgenome. *New Phytol* **208**, 4-8.

Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)* **30**, 923-930.

Ling, H.Q., Zhao, S., Liu, D., Wang, J., Sun, H., Zhang, C., Fan, H., Li, D., Dong, L., Tao, Y., Gao, C., Wu, H., Li, Y., Cui, Y., Guo, X., Zheng, S., Wang, B., Yu, K., Liang, Q., Yang, W., Lou, X., Chen, J., Feng, M., Jian, J., Zhang, X., Luo, G., Jiang, Y., Liu, J., Wang, Z., Sha, Y., Zhang, B., Wu, H., Tang, D., Shen, Q., Xue, P., Zou, S., Wang, X., Liu, X., Wang, F., Yang, Y., An, X., Dong, Z., Zhang, K., Zhang, X., Luo, M.C., Dvorak, J., Tong, Y., Wang, J., Yang, H., Li, Z., Wang, D., Zhang, A. and Wang, J. (2013) Draft genome of the wheat A-genome progenitor Triticum urartu. *Nature* **496**, 87-90.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H. and Ecker, J.R. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**, 523-536.

Lowe, R., Shirley, N., Bleackley, M., Dolan, S. and Shafee, T. (2017a) Transcriptomics technologies. *PLoS Comput Biol* **13**, e1005457.

Lowe, R., Shirley, N., Bleackley, M., Dolan, S. and Shafee, T. (2017b) Transcriptomics technologies. *PLOS Computational Biology* **13**, e1005457.

Luo, M.C., Gu, Y.Q., Puiu, D., Wang, H., Twardziok, S.O., Deal, K.R., Huo, N., Zhu, T., Wang, L., Wang, Y., McGuire, P.E., Liu, S., Long, H., Ramasamy, R.K., Rodriguez, J.C., Van,

S.L., Yuan, L., Wang, Z., Xia, Z., Xiao, L., Anderson, O.D., Ouyang, S., Liang, Y., Zimin, A.V., Pertea, G., Qi, P., Bennetzen, J.L., Dai, X., Dawson, M.W., Muller, H.G., Kugler, K., Rivarola-Duarte, L., Spannagl, M., Mayer, K.F.X., Lu, F.H., Bevan, M.W., Leroy, P., Li, P., You, F.M., Sun, Q., Liu, Z., Lyons, E., Wicker, T., Salzberg, S.L., Devos, K.M. and Dvorak, J. (2017) Genome sequence of the progenitor of the wheat D genome Aegilops tauschii. *Nature* **551**, 498-502.

Maere, S., Heymans, K. and Kuiper, M. (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics (Oxford, England)* **21**, 3448-3449.

Mago, R., Miah, H., Lawrence, G., Wellings, C., Spielmeyer, W., Bariana, H., McIntosh, R., Pryor, A. and Ellis, J. (2005) High-resolution mapping and mutation analysis separate the rust resistance genes Sr31, Lr26 and Yr9 on the short arm of rye chromosome 1. *Theoretical and Applied Genetics* **112**, 41-50.

Marcussen, T., Sandve, S.R., Heier, L., Spannagl, M., Pfeifer, M., International Wheat Genome Sequencing, C., Jakobsen, K.S., Wulff, B.B., Steuernagel, B., Mayer, K.F. and Olsen, O.A. (2014) Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**, 1250092.

Marra, M.A., Hillier, L. and Waterston, R.H. (1998) Expressed sequence tags--ESTablishing bridges between genomes. *Trends Genet* **14**, 4-7.

Martin, J.A. and Wang, Z. (2011) Next-generation transcriptome assembly. *Nat Rev Genet* **12**, 671-682.

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T., Radchuk, V., Dockter, C., Hedley, P.E., Russell, J., Bayer, M., Ramsay, L., Liu, H., Haberer, G., Zhang, X.Q., Zhang, Q., Barrero, R.A., Li, L., Taudien, S., Groth, M., Felder, M., Hastie, A., Simkova, H., Stankova, H., Vrana, J., Chan, S., Munoz-Amatriain, M., Ounit, R., Wanamaker, S., Bolser, D., Colmsee, C., Schmutzer, T., Aliyeva-Schnorr, L., Grasso, S., Tanskanen, J., Chailyan, A., Sampath, D., Heavens, D., Clissold, L., Cao, S., Chapman, B., Dai, F., Han, Y., Li, H., Li, X., Lin, C., McCooke, J.K., Tan, C., Wang, P., Wang, S., Yin, S., Zhou, G., Poland, J.A., Bellgard, M.I., Borisjuk, L., Houben, A., Dolezel, J., Ayling, S., Lonardi, S., Kersey, P., Langridge, P., Muehlbauer, G.J., Clark, M.D., Caccamo, M., Schulman, A.H., Mayer, K.F.X., Platzer, M., Close, T.J., Scholz, U., Hansson, M., Zhang, G., Braumann, I., Spannagl, M., Li, C., Waugh, R. and Stein, N. (2017) A chromosome conformation capture ordered sequence of the barley genome. *Nature* **544**, 427-433.

Mascher, M., Schuenemann, V.J., Davidovich, U., Marom, N., Himmelbach, A., Hubner, S., Korol, A., David, M., Reiter, E., Riehl, S., Schreiber, M., Vohr, S.H., Green, R.E., Dawson, I.K., Russell, J., Kilian, B., Muehlbauer, G.J., Waugh, R., Fahima, T., Krause, J., Weiss, E. and Stein, N. (2016) Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nat Genet* **48**, 1089-1093.

Mayer, K.F., Martis, M., Hedley, P.E., Simkova, H., Liu, H., Morris, J.A., Steuernagel, B., Taudien, S., Roessner, S., Gundlach, H., Kubalakova, M., Suchankova, P., Murat, F., Felder, M., Nussbaumer, T., Graner, A., Salse, J., Endo, T., Sakai, H., Tanaka, T., Itoh, T., Sato, K., Platzer, M., Matsumoto, T., Scholz, U., Dolezel, J., Waugh, R. and Stein, N. (2011a) Unlocking the barley genome by chromosomal and comparative genomics. *The Plant cell* **23**, 1249-1263.

Mayer, K.F.X., Martis, M., Hedley, P.E., Šimková, H., Liu, H., Morris, J.A., Steuernagel, B., Taudien, S., Roessner, S., Gundlach, H., Kubaláková, M., Suchánková, P., Murat, F., Felder, M., Nussbaumer, T., Graner, A., Salse, J., Endo, T., Sakai, H., Tanaka, T., Itoh, T., Sato, K., Platzer, M., Matsumoto, T., Scholz, U., Doležel, J., Waugh, R. and Stein, N. (2011b) Unlocking the Barley Genome by Chromosomal and Comparative Genomics. *The Plant cell* **23**, 1249-1263.

McIntosh, R. (1991) Alien sources of disease resistance in bread wheats. In: *Proceedings of Dr. H. Kihara Memorial International Symposium on Cytoplasmic Engineering in Wheat. Nuclear and Organellar Genomes of Wheat Species'.(Eds T Sasakuma, T Kinoshita) pp* pp. 320-332.

Mickelson-Young, L., Endo, T. and Gill, B. (1995) A cytogenetic ladder-map of the wheat homoeologous group-4 chromosomes. *Theoretical and Applied Genetics* **90**, 1007-1011.
Miftahudin, K., Ross, X., Ma, A. and Mahmoud, J. Layton et al., 2004 Analysis of expressed sequence tag loci on wheat chromosome group 4. *Genetics* **168**, 651-663.
Molnar-Lang, M., Kruppa, K., Cseh, A., Bucsi, J. and Linc, G. (2012) Identification and phenotypic description of new wheat: six-rowed winter barley disomic additions. *Genome* **55**, 302-311.
Molnar-Lang, M., Linc, G. and Szakacs, E. (2014) Wheat-barley hybridization: the last 40 years. *Euphytica* **195**, 315-329.
Mondal, S., Rutkoski, J.E., Velu, G., Singh, P.K., Crespo-Herrera, L.A., Guzmán, C., Bhavani, S., Lan, C., He, X. and Singh, R.P. (2016) Harnessing Diversity in Wheat to Enhance Grain Yield, Climate Resilience, Disease and Insect Pest Resistance and Nutrition Through Conventional and Modern Breeding Approaches. *Frontiers in plant science* **7**.
Moore, G., Devos, K.M., Wang, Z. and Gale, M.D. (1995) Cereal genome evolution. Grasses, line up and form a circle. *Current biology : CB* **5**, 737-739.
Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628.
Mujeeb-Kazi, A. (1995) Intergeneric crosses: Hybrid production and utilization. Utilizing Wild Grass Biodiversity in Wheat Improvement: 15 Years of Wide Cross Research at CIMMYT.
Murai, K., Koba, T. and Shimada, T. (1997) Effects of barley chromosome on heading characters in wheat-barley chromosome addition lines. *Euphytica* **96**, 281-287.
Murashige, T. and Skoog, F. (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum* **15**, 473-497.
Murat, F., Pont, C. and Salse, J. (2014) Paleogenomics in Triticeae for translational research. *Current Plant Biology* **1**, 34-39.
Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. and Snyder, M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344-1349.
Nelson, J.C., Sorrells, M.E., Van Deynze, A., Lu, Y.H., Atkinson, M., Bernard, M., Leroy, P., Faris, J.D. and Anderson, J.A. (1995) Molecular mapping of wheat: major genes and rearrangements in homoeologous groups 4, 5, and 7. *Genetics* **141**, 721-731.
Nevo, E. (1992) Origin, evolution, population genetics and resources for breeding of wild barley, Hordeum spontaneum, in the Fertile Crescent. *Barley: genetics, biochemistry, molecular biology and biotechnology.*, 19-43.
Nicolae, M., Mangul, S., Mandoiu, II and Zelikovsky, A. (2011) Estimation of alternative splicing isoform frequencies from RNA-Seq data. *Algorithms Mol Biol* **6**, 9.
Orabi, J., Backes, G., Wolday, A., Yahyaoui, A. and Jahoor, A. (2007) The Horn of Africa as a centre of barley diversification and a potential domestication site. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **114**, 1117-1127.
Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboob ur, R., Ware, D., Westhoff, P., Mayer, K.F., Messing, J. and Rokhsar, D.S. (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* **457**, 551-556.
Patro, R., Mount, S.M. and Kingsford, C. (2014) Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nature biotechnology* **32**, 462.
Petersen, G., Seberg, O., Yde, M. and Berthelsen, K. (2006) Phylogenetic relationships of Triticum and Aegilops and evidence for the origin of the A, B, and D genomes of common wheat (Triticum aestivum). *Mol Phylogenet Evol* **39**, 70-82.

Philippe, N., Salson, M., Commes, T. and Rivals, E. (2013) CRAC: an integrated approach to the analysis of RNA-seq reads. *Genome biology* **14**, R30.
Pingault, L., Choulet, F., Alberti, A., Glover, N., Wincker, P., Feuillet, C. and Paux, E. (2015) Deep transcriptome sequencing provides new insights into the structural and functional organization of the wheat genome. *Genome biology* **16**, 29.
Pose, D., Verhage, L., Ott, F., Yant, L., Mathieu, J., Angenent, G.C., Immink, R.G. and Schmid, M. (2013) Temperature-dependent regulation of flowering by antagonistic FLM variants. *Nature* **503**, 414-417.
Quelle, D.E., Zindy, F., Ashmun, R.A. and Sherr, C.J. (1995) Alternative Reading Frames of the Ink4a Tumor-Suppressor Gene Encode 2 Unrelated Proteins Capable of Inducing Cell-Cycle Arrest. *Cell* **83**, 993-1000.
Rabinovich, S. (1998) Importance of wheat-rye translocations for breeding modern cultivar of Triticum aestivum L. *Euphytica* **100**, 323-340.
Rasheed, A., Mujeeb-Kazi, A., Ogbonnaya, F.C., He, Z. and Rajaram, S. (2017) Wheat genetic resources in the post-genomics era: promise and challenges. *Ann Bot*.
Ray, D.K., Mueller, N.D., West, P.C. and Foley, J.A. (2013) Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PloS one* **8**, e66428.
Rayburn, A. and Gill, B. (1985) Use of biotin-labeled probes to map specific DNA sequences on wheat chromosomes. *Journal of Heredity* **76**, 78-81.
Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R. and Ruvkun, G. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, 901-906.
Rey, E., Molnár, I. and Doležel, J. (2015) Genomics of Wild Relatives and Alien Introgressions. In: *Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and Genomics* (Molnár-Láng, M., Ceoloni, C. and Doležel, J. eds), pp. 347-381. Cham: Springer International Publishing.
Reynolds, M., Calderini, D., Condon, A. and Rajaram, S. (2001) Physiological basis of yield gains in wheat associated with the LR19 translocation from Agropyron elongatum. In: *Wheat in a Global Environment* pp. 345-351. Springer.
Roberts, A. and Pachter, L. (2013) Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Methods* **10**, 71-73.
Roberts, A., Pimentel, H., Trapnell, C. and Pachter, L. (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics (Oxford, England)* **27**, 2325-2329.
Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y.S., Newsome, R., Chan, S.K., She, R., Varhol, R., Kamoh, B., Prabhu, A.L., Tam, A., Zhao, Y., Moore, R.A., Hirst, M., Marra, M.A., Jones, S.J., Hoodless, P.A. and Birol, I. (2010) De novo assembly and analysis of RNA-seq data. *Nat Methods* **7**, 909-912.
Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140.
Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology* **11**, R25.
Rustenholz, C., Choulet, F., Laugier, C., Safar, J., Simkova, H., Dolezel, J., Magni, F., Scalabrin, S., Cattonaro, F., Vautrin, S., Bellec, A., Berges, H., Feuillet, C. and Paux, E. (2011) A 3,000-loci transcription map of chromosome 3B unravels the structural and functional features of gene islands in hexaploid wheat. *Plant Physiology* **157**, 1596-1608.
Saisho, D. and Purugganan, M.D. (2007) Molecular phylogeography of domesticated barley traces expansion of agriculture in the Old World. *Genetics* **177**, 1765-1776.
Sandve, S.R., Marcussen, T., Mayer, K., Jakobsen, K.S., Heier, L., Steuernagel, B., Wulff, B.B. and Olsen, O.A. (2015) Chloroplast phylogeny of Triticum/Aegilops species is not incongruent with an ancient homoploid hybrid origin of the ancestor of the bread wheat D-genome. *New Phytol* **208**, 9-10.

Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-470.

Schnable, J.C., Freeling, M. and Lyons, E. (2012) Genome-wide analysis of syntenic gene deletion in the grasses. *Genome biology and evolution* **4**, 265-277.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P. and Reily, A.D. (2009a) The B73 maize genome: complexity, diversity, and dynamics. **326**, 1112-1115.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., Van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.M., Deragon, J.M., Estill, J.C., Fu, Y., Jeddeloh, J.A., Han, Y., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A. and Wilson, R.K. (2009b) The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112-1115.

Schulte, D., Close, T.J., Graner, A., Langridge, P., Matsumoto, T., Muehlbauer, G., Sato, K., Schulman, A.H., Waugh, R., Wise, R.P. and Stein, N. (2009) The International Barley Sequencing Consortium-At the Threshold of Efficient Access to the Barley Genome. *Plant Physiology* **149**, 142-147.

Schulz, M.H., Zerbino, D.R., Vingron, M. and Birney, E. (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics (Oxford, England)* **28**, 1086-1092.

Sears, E.R. (1956) *The transfer of leaf-rust resistance from Aegilops umbellulata to wheat*.

Sears, E.R. (1978) The telocentric chromosomes of common wheat. In: *Proc. 5th Intl. Wheat Genet. Symp., 1978* pp. 389-407. Indian Society of Genetics & Plant Breeding.

Shang, X., Cao, Y. and Ma, L. (2017) Alternative Splicing in Plant Genes: A Means of Regulating the Environmental Fitness of Plants. *Int J Mol Sci* **18**.

Sharma, S., Xu, S., Ehdaie, B., Hoops, A., Close, T.J., Lukaszewski, A.J. and Waines, J.G. (2011) Dissection of QTL effects for root traits using a chromosome arm-specific mapping population in bread wheat. *Theoretical and Applied Genetics* **122**, 759-769.

Staiger, D. and Brown, J.W.S. (2013) Alternative Splicing at the Intersection of Biological Timing, Development, and Stress Responses. *The Plant cell* **25**, 3640-3656.

Stirnweis, D., Milani, S.D., Brunner, S., Herren, G., Buchmann, G., Peditto, D., Jordan, T. and Keller, B. (2014) Suppression among alleles encoding nucleotide-binding-leucine-rich repeat resistance proteins interferes with resistance in F1 hybrid and allele-pyramided wheat plants. *The Plant journal : for cell and molecular biology* **79**, 893-903.

Stoces, S., Ruttink, T., Bartos, J., Studer, B., Yates, S., Zwierzykowski, Z., Abrouk, M., Roldan-Ruiz, I., Ksiazczyk, T., Rey, E., Dolezel, J. and Kopecky, D. (2016) Orthology Guided

Transcriptome Assembly of Italian Ryegrass and Meadow Fescue for Single-Nucleotide Polymorphism Discovery. *Plant Genome* **9**.

Szakacs, E. and Molnar-Lang, M. (2007) Development and molecular cytogenetic identification of new winter wheat--winter barley ('Martonvasari 9 kr1' - 'Igri') disomic addition lines. *Genome* **50**, 43-50.

Takahara, T., Kanazu, S.I., Yanagisawa, S. and Akanuma, H. (2000) Heterogeneous Sp1 mRNAs in human HepG2 cells include a product of homotypic trans-splicing. *J Biol Chem* **275**, 38067-38072.

Tang, J., Ohyama, K., Kawaura, K., Hashinokuchi, H., Kamiya, Y., Suzuki, M., Muranaka, T. and Ogihara, Y. (2011) A new insight into application for barley chromosome addition lines of common wheat: achievement of stigmasterol accumulation. *Plant Physiology* **157**, 1555-1567.

Tilman, D., Balzer, C., Hill, J. and Befort, B.L. (2011) Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences* **108**, 20260-20264.

Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)* **25**, 1105-1111.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. and Pachter, L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-578.

Turkosi, E., Cseh, A., Darko, E. and Molnar-Lang, M. (2016) Addition of Manas barley chromosome arms to the hexaploid wheat genome. *BMC Genet* **17**, 87.

van Ginkel, M. and Ogbonnaya, F. (2007) Using synthetic wheats to breed cultivars better adapted to changing production conditions. *Field Crops Research* **104**, 86-94.

Wang, K., Singh, D., Zeng, Z., Coleman, S.J., Huang, Y., Savich, G.L., He, X., Mieczkowski, P., Grimm, S.A., Perou, C.M., MacLeod, J.N., Chiang, D.Y., Prins, J.F. and Liu, J. (2010) MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res* **38**, e178.

Wang, X., Zhang, H., Li, Y., Zhang, Z., Li, L. and Liu, B. (2016) Transcriptome asymmetry in synthetic and natural allotetraploid wheats, revealed by RNA-sequencing. *New Phytol* **209**, 1264-1277.

Wang, Z., Gerstein, M. and Snyder, M. (2009a) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63.

Wang, Z., Gerstein, M. and Snyder, M. (2009b) RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics* **10**, 57.

Warburton, M., Crossa, J., Franco, J., Kazi, M., Trethowan, R., Rajaram, S., Pfeiffer, W., Zhang, P., Dreisigacker, S. and Van Ginkel, M. (2006) Bringing wild relatives back into the family: recovering genetic diversity in CIMMYT improved wheat germplasm. *Euphytica* **149**, 289-301.

Waterhouse, P.M., Graham, H.W. and Wang, M.B. (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *P Natl Acad Sci USA* **95**, 13959-13964.

Wilson, S. (1873) II . Wheat and Rye Hybrids. *Transactions of the Botanical Society of Edinburgh* **12**, 286-288.

Woodhouse, M.R., Schnable, J.C., Pedersen, B.S., Lyons, E., Lisch, D., Subramaniam, S. and Freeling, M. (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. *PLoS biology* **8**, e1000409.

Wu, T.D. and Nacu, S. (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics (Oxford, England)* **26**, 873-881.

Wu, Y., Sun, Y., Wang, X., Lin, X., Sun, S., Shen, K., Wang, J., Jiang, T., Zhong, S., Xu, C. and Liu, B. (2016) Transcriptome shock in an interspecific F1 triploid hybrid of Oryza revealed by RNA sequencing. *J Integr Plant Biol* **58**, 150-164.

Wulff, B.B. and Moscou, M.J. (2014) Strategies for transferring resistance into wheat: from wide crosses to GM cassettes. *Frontiers in plant science* **5**, 692.

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T.W., Li, Y., Xu, X., Wong, G.K. and Wang, J. (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics (Oxford, England)* **30**, 1660-1666.

Yamamoto, M. and Mukai, Y. (1989) Application of fluorescence in situ hybridization to molecular cytogenetics of wheat. *Wheat Information Service*, 30-32.

Zarco-Hernandez, J.A., Santiveri, F., Michelena, A. and Peña, R.J. (2005) Durum wheat (Triticum turgidum, L.) carrying the 1BL/1RS chromosomal translocation: agronomic performance and quality characteristics under Mediterranean conditions. *European journal of agronomy* **22**, 33-43.

Zeng, W. and Mortazavi, A. (2012) Technical considerations for functional sequencing assays. *Nature Immunology* **13**, 802.

Zhang, J.F., Li, G.R., Li, H.J., Pu, X.B., Jiang, J., Chai, L., Zheng, B.C., Cui, C., Yang, Z.J., Zhu, Y.Q. and Jiang, L.C. (2015) Transcriptome Analysis of Interspecific Hybrid between Brassica napus and B. rapa Reveals Heterosis for Oil Rape Improvement. *Int J Genomics*.

Zhang, Z. and Wang, W. (2014) RNA-Skim: a rapid method for RNA-Seq quantification at transcript level. *Bioinformatics (Oxford, England)* **30**, i283-i292.

Zimin, A.V., Puiu, D., Hall, R., Kingan, S., Clavijo, B.J. and Salzberg, S.L. (2017) The first near-complete assembly of the hexaploid bread wheat genome, Triticum aestivum. *Gigascience* **6**, 1-7.

3 Objectives and results

3.1 Objectives of the Thesis

Wide hybridization has been shown to be a successful strategy to (re)introduce novel genetic variation for various agronomical traits underlying better adaption to adverse environment conditions and resistance to biotic stresses into the bread wheat elite germplasm. However, relatively few wheat-alien introgression lines have made their way to agricultural use due to negative effects of post-hybridization interactions between the host genome and alien chromatin on the functionality of the host genome and therefore on agronomical performance. Understanding how the host and alien genomes interact at various functional levels (i.e. genomic, transcriptomic, proteomic, epigenetic, …) is necessary to make the development of useful introgression lines more efficient and successful. For a long time, the lack of genomic resources prevented detailed whole genome functional analyses. However, the recent release of reference genome sequences of hexaploid wheat and barley changed the paradigm. During the course of this work we made the use of these novel sequence resources combined with up-to-date bioinformatics technologies to investigate transcriptomic changes affecting host and alien genes as a result of interactions between the two genomes in a wheat-barley chromosome-arm addition line, which was chosen as a model system because of its potential for wheat improvement.

The general aim of this work comprised two main objectives:

3.1.1 Establish a protocol for transcriptome analysis in wheat-barley alien introgression lines

Because transcriptomic analysis of a wheat-alien introgression line was never performed at a genome-wide level, the first main objective of the Thesis was to define the best technological platform, plant material, experimental design and transcriptome analysis pipeline for the transcriptomic analysis.

3.1.2 Study the impact of inter-genome interactions on gene expression in wheat-barley 7HL introgression line

Once accurate and sensitive transcriptomic changes were assessed for both wheat and barley genes in the introgression plant material, the second objective of the Thesis was to analyse the changes in gene transcription within their genomic, evolutionary and synteny context in order to better understand by which mechanisms they were affected.

3.2 Results of the Thesis

3.2.1 Original papers

3.2.1.1 First author book chapter: "Genomics of wild relatives and Alien Introgressions". (Appendix I)

3.2.1.2 Peer reviewed first author paper: "Transcriptome reprogramming due to the introduction of a barley telosome into bread wheat affects more barley genes than wheat". (Appendix II)

3.2.1.1 Genomics of wild relatives and Alien Introgressions

Elodie Rey, István Molnár, Jaroslav Doležel

Book Chapter

In: *Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and Genomics* (Molnár-Láng, M., Ceoloni, C. and Doležel, J. eds), pp. 347-381. Cham: Springer International Publishing.

DOI: https://doi.org/10.1007/978-3-319-23494-6_13

Abstract:

Alien introgression breeding is an attractive approach to recover genetic variation that was lost during wheat domestication and breeding. New alleles and genes may be introduced from wild relatives from the tribe Triticeae, which exhibits large genetic variation and many potentially useful traits. Although a range of wheat–alien introgression lines has been developed, apart from the 1BL.1RS translocation, only a few commercial wheat cultivars benefitted from alien introgression. This is a consequence of poor knowledge of genome structure of wild donors, limited ability to control chromosome behavior during meiosis in interspecific hybrids and introgression lines, difficulties in eliminating unwanted chromatin transferred simultaneously with genes of interest, as well as a lack of tools permitting large-scale production and characterization of introgression lines. Recent advances in molecular and flow cytogenetics and genomics are bound to change the situation. New insights into the meiotic recombination raise the hopes for the ability to control its frequency and distribution. The progress in comparative genome analysis provides clues about the genome collinearity between wild species and wheat, making it possible to assess chances for chromosome recombination and predict its outcomes. Genomics tools enable massive and high-resolution screening of hybrids and their progenies and characterize their genomes, including the development of markers linked to traits of interest. Until recently, little attention has been paid to the function of introgressed genes and their interaction with the host genome. However, this has been changing and all these

achievements make the breeding of improved wheat cultivars using wild germplasm a realistic goal.

3.2.1.2 Transcriptome reprogramming due to the introduction of a barley telosome into bread wheat affects more barley genes than wheat

Elodie Rey, Michael Abrouk, Gabriel Keeble-Gagnère, Miroslava Karafiátová, Jan Vrána, Sandrine Balzergue, Ludivine Soubigou-Taconnat, Véronique Brunaud, Marie-Laure Martin-Magniette, Takashi R. Endo, Jan Bartoš, IWGSC, Rudi Appels and Jaroslav Doležel

Accepted paper

Plant Biotechnology Journal

DOI: 10.1111/pbi.12913.

Abstract:

Despite a long history, the production of useful alien introgression lines in wheat remains difficult mainly due to linkage drag and incomplete genetic compensation. In addition, little is known about the molecular mechanisms underlying the impact of foreign chromatin on plant phenotype. Here, a comparison of the transcriptomes of barley, wheat and a wheat barley 7HL addition line allowed the transcriptional impact both on 7HL genes of a non-native genetic background, and on the wheat gene complement as a result of the presence of 7HL to be assessed. Some 42% (389/923) of the 7HL genes assayed were differentially transcribed, which was the case for only 3% (960/35,301) of the wheat gene complement. The absence of any transcript in the addition line of a suite of chromosome 7A genes implied the presence of a 36 Mbp deletion at the distal end of the 7AL arm; this deletion was found to be in common across the full set of Chinese Spring/Betzes barley addition lines. The remaining differentially transcribed wheat genes were distributed across the whole genome. The up-regulated barley genes were mostly located in the proximal part of the 7HL arm, while the down-regulated ones were concentrated in the distal part; as a result, genes encoding basal cellular functions tended to be transcribed, while those encoding specific functions were suppressed. An insight has been gained into gene transcription in an alien introgression line, thereby providing a basis for understanding the interactions between wheat and exotic genes in introgression materials.

3.2.2 Published abstracts – poster presentations

3.2.2.1 ***"Transcriptomic analysis of interactions between host and alien genomes in wheat-barley interspecific introgression" presented at the International Conference on Genomics and Bioinformatics, Izmir (May 7th 2016) (Appendix III)***

3.2.2.2 ***"Genomics of alien gene transfer in wheat with an emphasis on the Transcriptomic analysis performed on barley 7HL genes" presented at the European Wheat Aneuploid Conference (EWAC), Lublin (May 25-29 2015) (Appendix IV)***

3.2.2.3 ***"Genomics of alien gene transfer in wheat" presented at the International ITMI-Eucarpia conference on cereals, Wernigerode (June 29-July 4, 2014) (Appendix V)***

3.2.2.1 Transcriptomic analysis of interactions between host and alien genomes in wheat-barley interspecific introgression

Elodie Rey, Gabriel Keeble-Gagnère, Sandrine Balzergue, Ludivine Soubigou-Taconnat, Véronique Brunaud, Marie-Laure Martin-Magniette, Takashi R. Endo, Michael Abrouk, Jan Bartoš, Rudi Appels and Jaroslav Doležel

In abstracts of: "International Conference on Genomics and Bioinformatics". Izmir, Turkey, 2016.

For wheat as for many other crops, it is mandatory to increase genetic diversity of current bread varieties in order to satisfy the demands of the growing human population and sustain the productivity threatened by changing climate. An attractive approach to recover genetic variation is to cross wheat with its relatives from the tribe Triticeae to introgress valuable traits for agronomical improvement, a process called alien introgression. Despite being practiced for more than a century, alien introgression in wheat remains a tedious process impeded by pre- and post-hybridization incompatibility barriers. The aim of this project is to provide novel information on the interactions between a host genome and the introgressed alien chromatin, both at the genome and transcriptome levels. We have chosen wheat-barley ditelosomic addition line 7HL, which was obtained after hybridization between hexaploid wheat cv. Chinese Spring ($2n=6x=42$, AABBDD) and barley cv. Betzes ($2n=2x=14$, HH), as a model system. The introgressed material contains a complete set of wheat chromosomes and a pair of long arms of barley chromosome 7H, and exhibits interesting agronomic traits such as salt tolerance, earliness and increased β-glucan content in the grain. We have sequenced mRNA of both parents and the addition line 7HL to reveal potential changes in the patterns of expression resulting from the intergenomic interactions. As a consequence of the hexaploid nature of wheat genome and its relatedness with barley genome, one of the difficulties in our analysis is a reliable discrimination between the homoeologous transcripts of the wheat genome and barley transcripts in the ditelosomic line. Therefore, as a first step, we had to evaluate the degree of homology between wheat and barley

genic sequences in order to set the best parameters for the critical step of mapping reads against the assembly of transcripts before differential gene expression (DE) analysis. The results of a BLAST analysis of wheat CDS against barley CDS allowed us to identify a set of nearly 400 genes in wheat and barley for which their sequences share 90% identity over the complete alignment length. This result led us to increase the stringency of the mapping of the reads against the transcripts, and allowed us to check the expression for these problematic genes. Then, the DE analysis was performed by comparing the expression of wheat and barley genes in the natural background with the expression of the same gene set in the introgression line. This allowed us to observe that most (98.7%) alien genes introgressed into wheat genetic background are repressed, while host genes are mostly (98.5%) non-affected in their expression. Interestingly, a small proportion of barley 7HL genes (1.1%) are not differentially expressed in wheat background compared to in the barley parent, and few wheat genes are up- (0.5%) or down- (0.9%) regulated in the presence of alien chromatin. The analysis of those genes within their genomic context will allow us to infer the effect of structural and functional features on the changes in gene expression observed, which might lead us to understand how the changes in gene expression are directed, and if they do play a role in the weak phenotype of these introgression lines. The outputs from this work include suites of genes that may relate to the process of making inter-species crosses more compatible so that more wheat-alien introgression lines will make their way into practical use.

3.2.2.2 Genomics of alien gene transfer in wheat with an emphasis on the Transcriptomic analysis performed on barley 7HL genes

Elodie Rey, Sandrine Balzergue, Ludivine Soubigou-Taconnat, Véronique Brunaud, Endo R. Takashi, Michael Abrouk, Jan Bartos and Jaroslav Dolezel

In abstracts of: "European Wheat Aneuploid Conference (EWAC) 2015 International Conference". Lublin, Poland, 2015.

Since more than a century, wheat has been crossed with its wild relatives in attempt to introduce favourable alleles, genes and gene complexes providing resistance to biotic and abiotic stresses and other beneficial traits. Although hybridization with wild species represents a powerful approach for crop improvement, the knowledge of underlying biological mechanisms remains limited and only a few wheat-alien introgression lines have made their way into the agriculture. This project aims to provide novel information on the interaction between a host genome and introgressed chromosomes and their parts, both at genome and transcriptome levels. We have chosen wheat-barley ditelosomic addition line 7HL, which was obtained after hybridization between bread wheat cv. Chinese Spring and barley cv. Betzes, as a model system. The addition line 7HL contains a complete set of wheat chromosomes and a pair of long arms of barley chromosome 7H, and exhibits interesting agronomic traits such as salt tolerance, earliness and increased β-glucan content in the grain. We have sequenced transcriptomes of both parents and the ditelosomic addition line 7HL to reveal differences in gene expression due to the intergenomic interactions. As a first step, we compared expression of 7HL genes in the natural background (in barley parent) with the expression of the same gene set in the wheat host background. While a majority of 7HL genes were down-regulated in the host background, some genes were up-regulated. This analysis provides new data to understand the regulation of alien genes in a recipient genome.

3.2.2.3 Genomics of alien gene transfer in wheat

Elodie Rey, Sandrine Balzergue, Ludivine Soubigou-Taconnat, Véronique Brunaud, Endo R. Takashi, Jan Bartos and Jaroslav Dolezel

In abstracts of: "EUCARPIA Cereals Section – I T M I Joint Conference".
Wernigerode, Germany, 2014.

Wheat, maize and rice are the three most important crops worldwide. Wheat itself is feeding nearly half of the world's population (40%) and provides 20% of the total food calories and proteins in human nutrition. There is a big concern that productivity and quality of cultivated wheat continue to increase in order to satisfy the need for food by the growing human population. One of the strategies for wheat improvement is to enrich the existing gene pool of cultivated wheat by introducing favourable alleles, genes and gene complexes from wild and close relatives. Although alien gene introgression represents a powerful approach for crop improvement, the knowledge of underlying biological mechanisms remains limited. This project aims to provide novel information on the interaction between a host genome and introgressed chromosomes and their parts, both at genome structure and transcriptome levels. We have chosen wheat-barley ditelosomic addition line 7HL as a model system. The line was obtained after interspecific hybridization between wheat cv. Chinese Spring and barley cv. Betzes. The addition line 7HL contains the full set of wheat chromosomes and a pair of long arms of barley chromosome 7H, and exhibits interesting agronomic traits such as salt tolerance, earliness and high β-glucan content. We have sequenced transcriptomes of both parental species and the addition line 7HL to compare gene expression between the ditelosomic line and the parents. One of the difficulties is to discriminate homoeologous transcripts from duplicated copies in the wheat genome, and to distinguish barley transcripts from their host relatives in the ditelosomic line background. The results of this study should provide new information to understand the regulation of the alien genes in the host genome, and reciprocally the impact of the additional alien genes copies on the expression of their host homoeologues.

3.2.3 Published abstracts – oral presentations

3.2.3.1 "The impact of genomic interactions on gene expression in a wheat-barley 7HL addition line" presented at the International conference PAG XXV, San Diego (14-18 Jan. 2017)

3.2.3.2 "A transcriptomic analysis of wheat-barley 7HL-addition line: Genomic interactions between host and alien genomes in products of interspecific hybridization" presented at International conference PAG ASIA 2016, Singapore (6–9 Jun. 2016)

3.2.3.1 The impact of genomic interactions on gene expression in a wheat-barley 7HL addition line

Elodie Rey, Gabriel Keeble-Gagnère, Sandrine Balzergue, Ludivine Soubigou-Taconnat, Véronique Brunaud, Marie-Laure Martin-Magniette, Takashi R. Endo, Michael Abrouk, Jan Bartoš, Rudi Appels and Jaroslav Doležel

In abstract of: "The Plant and Animal Genome XXVI Conference (PAG)".

San Diego, USA, 2017

Alien gene transfer has been practiced for more than a century as a way to improve wheat by introducing agronomic favourable genes or alleles from its relatives, wild or cultivated. However, only few wheat-alien introgressions have made their way into practical use in agriculture, due to genetic incompatibility between wheat and its relatives. Despite a keen interest in transferring favourable agronomical traits from barley (earliness, salt tolerance and drought tolerance) into wheat, very few wheat-barley introgression lines have been reported. In order to characterize changes in the function of a gene transferred into a different genome, we performed a RNA-seq analysis using a wheat-barley ditelosomic addition line 7HL, together with the host hexaploid wheat cv. Chinese Spring and the 7HL donor barley cv. Betzes. Comparison of the expression of barley genes in their own genetic background and in wheat background revealed extensive down-regulation of the 7HL genes introgressed in the addition line. On the other hand, only a small fraction of the wheat host genes had altered expression in the addition line carrying the chromosome 7HL of barley, with 490 genes being down-regulated and 243 genes up-regulated. However, this small effective account with important components of central metabolic and stress related pathways for the plant, such as carbohydrate metabolism, photosynthesis, and heat-shock response. The analysis of the gene networks affected by the changes in expression of the wheat and barley genes in the ditelosomic addition line 7HL, provide a basis for understanding the effects of alien gene transfers into wheat.

3.2.3.2 A transcriptomic analysis of wheat-barley 7HL-addition line: Genomic interactions between host and alien genomes in products of interspecific hybridization

Elodie Rey, Gabriel Keeble-Gagnère, Sandrine Balzergue, Ludivine Soubigou-Taconnat, Véronique Brunaud, Marie-Laure Martin-Magniette, Takashi R. Endo, Michael Abrouk, Jan Bartoš, Rudi Appels and Jaroslav Doležel

In abstract of: "PAG ASIA 2016".

Singapore, 2016.

For wheat as for many other crops, it is mandatory to increase genetic diversity of current bread varieties in order to satisfy the demands of the growing human population and sustain the productivity threatened by changing climate. While alien gene transfer represents an attractive approach for wheat improvement practiced for nearly 150 years, this strategy still suffers from major drawbacks reflecting genetic incompatibility between wheat and its relatives. Here we performed an RNA-seq analysis of hexaploid wheat cv. Chinese Spring (2*n*=6*x*=42, AABBDD), barley cv. Betzes (2*n*=2*x*=14, HH) and a wheat^{CS}-barley^{Bet} ditelosomic addition line 7HL resulting from the interspecific hybridization between them, in order to reveal changes in the patterns of expression resulting from the intergenomic interactions. The comparison of the expression of wheat and barley genes in the natural background with the expression of the same gene set in the introgression line allowed us to observe that most (98.7%) alien genes introgressed into wheat genetic background are repressed, while host genes are mostly (98.5%) non-affected in their expression. Interestingly, a small proportion of barley 7HL genes (1.1%) are not differentially expressed in wheat background compared to in the barley parent, and few wheat genes are up- (0.5%) or down- (0.9%) regulated in the presence of alien chromatin. The analysis of those genes within their genomic, structural and evolutionary context allow us identifying genes that may relate to the process of making inter-species crosses more compatible so that more wheat-alien introgression lines make their way into practical use.

3.2.4 Awards

3.2.4.1 EMBO short-term fellowship (Award ASTF 485 – 2015) awarded to support three months research stay (Nov. 2015 – Feb. 2016) at School of bioscience of the University of Melbourne (AU) (Appendix VI)

4 Conclusions

In this thesis, I characterized changes in the expression of host and alien genes in wheat-barley chromosome-arm addition line 7HL. The study was motivated by a desire to understand the interactions between alien chromatin and host genome, which are critical for rational use of alien introgression breeding in wheat improvement programmes, and in other crops in general. The first goal was to develop a protocol for transcriptome analysis in wheat-barley alien introgression lines. Accurate assessment of the expression of wheat and barley genes in the context of the hexaploid nature of the wheat genome and its relatedness with barley genome is computationally and statistically demanding due to sequence homology between wheat and barley transcripts. Using the latest genome sequence resources for wheat and barley, a high level of biological replication and more stringent criteria than normally used for RNA-seq analysis, we were able for the first time to assess the expression of about 40% of wheat and barley annotated gene content in a wheat-barley introgression line. Our results showed that a large proportion (~42%) of genes on barley 7HL chromosome arm were differentially transcribed in the addition line relative to barley alone, while the expression of only 3% of wheat genes was affected. This pioneering study opens the way for further transcriptomic analysis in other alien introgression lines, including breeding plant material. The second part of the Thesis was dedicated to the study of changes in wheat host and barley introgressed gene expression within their genomic and syntenic context. Our analysis revealed a large deletion on wheat chromosome arm 7AL in the addition line. This was not expected and the deletion was subsequently identified in all addition lines developed from the wheat cv Chinese Spring / barley cv Betzes combination, and was likely present in the stock of wheat used to generate the disomic addition line set. Moreover, the genes experiencing differential transcription in the addition line appeared to have different functions. The present study identified a set of genes of potential relevance for determining gene networks responsible for interspecific incompatibility between wheat and its related species. This set of genes also provides a resource for investigating the molecular basis of differential transcription, which may

involve regulatory sequences, epigenetic changes or spatial organization of chromosome domains in the interphase nucleus.

5 List of abbreviations

1C	Amount of DNA in non-replicated haploid chromosome set
BC	Backcross
cDNA	Complementary DNA
CPM	Counts per million
CSS	Chromosome survey sequences
cv	Cultivar
DE	Differentially expressed
DGE	Differentially expressed genes
DNA	Deoxyribonucleic acid
EST	Expressed sequence tag
FISH	Fluorescence *in situ* hybridization
FPKM	Fragments per kilobase of transcripts per million mapped reads
Gb	Gigabase
GISH	Genomic *in situ* hybridization
GO	Gene onthology
IBSC	International Barley Sequencing Consortium
IWGSC	International Wheat Genome Sequencing Consortium
lncRNA	Long non-coding RNA
LTR	Long terminal repeat
Mb	Mega bases
miRNA	Micro RNA
mRNA	Messenger RNA
MTP	Minimal tilling path
MYA	Million years ago
ncRNA	Non-coding RNA

NGS	Next generation sequencing
ORF	Open reading frame
PE	Paired-end
pre-mRNA	Precursor mRNA
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RPKM	Reads per kilobase of transcripts per million mapped reads
rRNA	Ribosomal RNA
RT-qPCR	Reverse transcription quantitative PCR
SAGE	Serial analysis of gene expression
SE	Single end
siRNA	Small interfering RNA
snRNA	Small nuclear RNA
TPM	Transcript per million mapped reads
UTR	Untranslated region
YA	Years ago

6 List of appendices

Original papers

Appendix I: Genomics of wild relatives and Alien Introgressions

Appendix II: Transcriptome reprogramming due to the introduction of a barley telosome into bread wheat affects more barley genes than wheat

Published abstracts – poster presentations

Appendix III: Transcriptomic analysis of interactions between host and alien genomes in wheat-barley interspecific introgression

Appendix IV: Transcriptomic analysis of barley 7HL genes introgressed in wheat

Appendix V: Genomics of alien gene transfer in wheat

Awards

Appendix VI: Award ASTF 485 – 2015

6.1 Appendix I

Genomics of wild relatives and Alien Introgressions

Elodie Rey, István Molnár, Jaroslav Doležel

Book Chapter

In: Molnár-Láng M., Ceoloni C., Doležel J. - Alien Introgression in Wheat.

Springer, Cham

DOI: https://doi.org/10.1007/978-3-319-23494-6_13

Chapter 13
Genomics of Wild Relatives and Alien Introgressions

Elodie Rey, István Molnár, and Jaroslav Doležel

13.1 Introduction

As one of the most important staple food crops, bread wheat (*Triticum aestivum*, L.) continues to play a major role in ensuring global food security. The growing human population is estimated to reach nine billion by 2050, and in order to meet the expected demand, the annual yield increase of wheat should reach 2 %. This is a great challenge, as climate change and land degradation act against this endeavor. Apart from improved agronomic practice and reduction of postharvest losses, the key elements will be new varieties with increased resistance to diseases and pests, adverse environmental conditions, and with improved yield.

According to the most widely accepted scenario, bread wheat ($2n=6x=42$, BBAADD genome) arose about 8000 years ago when a cultivated form of tetraploid *Triticum turgidum* ($2n=4x=28$, BBAA genome) migrated to south of the Caspian Sea and in the area of Fertile Crescent crossed with a wild diploid grass *Aegilops tauschii* Coss. ($2n=2x=14$, DD genome). The union of unreduced gametes, or somatic chromosome doubling in the hybrid (Feuillet et al. 2008), resulted in a new allohexaploid species. The genetic diversity of bread wheat was restricted at the onset of its origin by the limited diversity of parental populations and was eroded subsequently during domestication and thousands years of cultivation and breeding.

One option to recover the useful variation that was lost and to acquire new and valuable genes and alleles is to utilize wild relatives of wheat, which were not

E. Rey • J. Doležel (✉)
Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, Olomouc 78371, Czech Republic
e-mail: dolezel@ueb.cas.cz

I. Molnár
Agricultural Institute, Centre for Agricultural Research, Hungarian Academy of Sciences, Brunszvik u. 2, Martonvásár 2462, Hungary

M. Molnár-Láng et al. (eds.), *Alien Introgression in Wheat*,
DOI 10.1007/978-3-319-23494-6_13

subjected to human selection, and thus represent a rich source of diversity. The tribe Triticeae comprises wild annual and perennial species related to wheat, facilitating the production of interspecific hybrids. The efforts to use this approach date back 140 years, and the first experiments at the end of nineteenth century and beginning of twentieth century involved hybridization between wheat and rye (Wilson 1876), wheat and barley (Farrer 1904), and between wheat and *Aegilops* (Kihara 1937). However, larger-scale production of interspecific hybrids was delayed until the introduction of colchicine treatment in 1930s (Blakeslee 1937), allowing the production of fertile amphiploids by doubling chromosome number in otherwise sterile hybrids. Among other, this provided a way to develop triticale as a new cereal crop (Meurant 1982). With the advances in hybridization techniques (Kruse 1973) and establishment of *in vitro* embryo rescue methodology (Murashige and Skoog 1962), wide hybridization became more accessible, and the experiments involved a larger group of wild and cultivated wheat relatives (Mujeeb-Kazi 1995).

An extensively used approach to utilize wild germplasm in wheat breeding has been the production of synthetic hexaploid wheat by hybridizing tetraploid durum wheat (*T. turgidum* ssp. *durum* (Desf.) Husn.) ($2n=4x=28$; BBAA genome) with *Ae. tauschii*. Both synthetic hexaploid and bread wheat have the same genomic constitution and therefore can be readily hybridized to transfer novel alleles and genes from different accessions of the D-genome progenitor. This strategy has been employed at CIMMYT where more than 1000 synthetic wheats were created (del Blanco et al. 2001; Warburton et al. 2006; van Ginkel and Ogbonnaya 2008; Li et al. 2014).

Genetic diversity suitable for wheat improvement is not limited to *Ae. tauschii*, and over the years, a range of interspecific hybrids, chromosome addition and translocation lines were obtained between perennial and annual Triticeae species and bread wheat (Mujeeb-Kazi 1995; Friebe et al. 1996; Schneider et al. 2008; Molnár-Láng et al. 2014). Probably the best example of a successful wheat–alien introgression has been the spontaneous 1BL.1RS chromosome translocation (Mujeeb-Kazi 1995). It was estimated that between 1991 and 1995, 45 % of 505 commercial cultivars of bread wheat in 17 countries carried 1BL.1RS translocation, which confers increased grain yield by providing race-specific disease resistance to major rust diseases (including *Lr29/Yr26* leaf and yellow rust resistance genes), improved adaptation and stress tolerance, superior aerial biomass, and higher kernel weight (Rabinovich 1998; Feuillet et al. 2008; Zarco-Hernandez et al. 2005). However, too few other alien introgressions into wheat made their way to agricultural practice.

This chapter reviews the progress in characterizing nuclear genomes of wild relatives of wheat and wheat–alien introgression lines at chromosomal and DNA levels, and the potential of these approaches to support wheat–alien introgression breeding. After introducing the diversity of wild relatives of wheat and the difficulties of the introgression breeding, methods of cytogenetics and genomics are outlined and examples of their uses are given. The need for better understanding the mechanisms controlling chromosome behavior and for better knowledge of genome structure of wild relatives is explained. The last part of the chapter is devoted to the interaction of the introgressed chromatin with the host wheat genome. This research area has been poorly developed so far, and the lack of information may hamper the attempts to develop improved cultivars of wheat with alien introgressions.

13.2 Wild Relatives of Wheat and Difficulties with Alien Introgression

The tribe Triticeae comprises a group of species belonging to the Poaceae grass family commonly named Gramineae. In addition to economically important bread wheat (*T. aestivum* L.), durum wheat (*Triticum turgidum* L. subsp. *durum (Desf.) Husn.*), barley (*Hordeum vulgare* L.), and rye (*Secale cereale* L.), the tribe comprises over 500 wild and cultivated species of genera *Aegilops*, *Agropyron*, *Ambylopyrum*, *Anthosachne*, *Campeiostachys*, *Dasypyrum*, *Elymus*, *Hordeum*, *Leymus*, *Lophopyrum*, *Psathyrostachys*, *Pseudoroegneria*, *Secale*, *Thinopyrum*, and *Triticum*.

The Triticeae species exhibit a large diversity in terms of geographical distribution, environmental requirements, and agronomically interesting traits. The latter includes increased yield (Reynolds et al. 2001), resistance to pests and diseases (Friebe et al. 1996), early maturity (Koba et al. 1997), drought tolerance (Fatih 1983; Molnár et al. 2004; Dulai et al. 2014), salt tolerance (Fatih 1983; Dulai et al. 2010; Darkó et al. 2015), micronutrient content and efficiency (Schlegel et al. 1998, Farkas et al. 2014), lodging resistance (Chen et al. 2012), heat tolerance (Pradhan and Prasad 2015), high dietary fibre content (Cseh et al. 2011), and high protein content (Pace et al. 2001). Donors for these traits have been identified and some of the traits have been transferred to wheat (Gill et al. 2011). Some of the genes responsible for the traits have been tagged, and a few of them were even cloned (Feuillet et al. 2008; Hajjar and Hodgkin 2007; Jiang et al. 1993). However, the degree of genetic and genomic characterization of wild Triticeae species is highly variable and uneven.

Although the potential of wild relatives for wheat improvement has been recognized since a long time, the available genetic diversity remains largely underexploited. In order to utilize its full potential, it is important to understand genome organization in wild wheat relatives, increase the number of genome-specific molecular tools and identify loci underlying traits of interest (Hajjar and Hodgkin 2007). The poor knowledge on genome structure of Triticeae species and the lack of high resolution genetic maps hampers identification of genes underlying important traits, identification of unwanted sequences and their elimination using suitable large-scale screening platforms.

Elimination of unwanted alleles may be challenging due to low level of recombination between chromosomes of wild relatives and wheat. Two principal approaches have been developed to overcome this hindrance. The first is based on decreasing the effect of *Ph1* locus by the use of wheat genotypes *ph1b* or *Ph*^{I} (Riley and Chapman 1958; Griffiths et al. 2006), which promotes recombination between homoeologous wheat and alien chromosomes. The second approach involves induction of donor chromosome breakage by ionizing irradiation, or gametocidal chromosomes (Jiang et al. 1993) to stimulate insertion of alien chromosome fragments into wheat chromosomes.

Evolutionary chromosome rearrangements broke down the collinearity between the homoeologous wheat and alien chromosomes (Devos and Gale 1993). As a consequence, genes on alien chromosome segments may not compensate for the loss of wheat genes. This may negatively affect agricultural performance of the wheat–alien introgression lines and represents another obstacle in using wheat–alien translocations

in breeding. Little is known about different levels of interaction between the host genome and the alien chromatin, which may lead to unexpected and even undesirable effects. Insertion of alien chromosome segment may interfere with functionality of the host genome at genomic, epigenomic, transcriptomic and proteomic levels, and may explain the failure of some introgressed genes to function in the host background, although their sequences remained intact after the introgression.

13.3 Tools to Support Alien Introgression in Wheat

13.3.1 Cytogenetics Techniques

The development of alien chromosome addition and translocation lines and their characterization greatly profits from the ability to identify chromosomes involved. Originally, the repertoire of selection methods was limited to cytological techniques that visualize mitotic and meiotic chromosomes. When Sears (1956) transferred leaf rust resistance from *Ae. umbellulata* to wheat, cytological characterization of the wheat—*Ae. umbellulata* addition line was limited to microscopic observation of mitotic chromosomes in root tips, and the translocation event was identified based on the leaf rust-resistance phenotype (Sears 1956). The advent of chromosome banding techniques such as Giemsa C-banding (Gill and Kimber 1974), permitted description of genomic constitution in interspecific hybrids, identification of alien chromosomes and characterization of translocations at subchromosomal level. C-banding was particularly effective in characterizing wheat–rye translocations because of diagnostic terminal bands of rye chromosomes (Lukaszewski and Gustafson 1983; Lapitan et al. 1984; Friebe and Larter 1988). However, it has been less useful if chromosomal segments of interest lacked diagnostic bands.

Introduction of techniques for *in situ* hybridization further stimulated the development and characterization of alien introgression lines. Following the pioneering work of Rayburn and Gill (1985), fluorescence *in situ* hybridization (FISH) was developed in wheat (Yamamoto and Mukai 1989). The potential of FISH to identify chromosomes and their segments depends on the availability of suitable probes. The most popular probes included the pAs1 repeat (Rayburn and Gill 1985; Nagaki et al. 1995), which permits identification of D-genome chromosomes, the rye subtelomeric repeat pSc119.2 (Bedbrook et al. 1980), which is useful to identify B-genome chromosomes, and pTa71 DNA clone (Gerlach and Bedbrook 1979), which identifies nucleolus organizing regions on satellite chromosomes. FISH with these probes discriminates the whole set of D- and B-genome chromosomes and, depending on the quality of hybridization, partially or completely the A-genome chromosomes of bread wheat. The same set of DNA probes has been applied to examine genetic diversity and construct karyotypes of wild species in *Aegilops* (Badaeva et al. 1996a, 1996b), *Agropyron* (Linc et al. 2012), and *Hordeum* (de Bustos et al. 1996; Szakács et al. 2013;), and to identify their chromosomes introgressed into wheat (Molnár et al. 2009; Sepsi et al. 2008; Nagy et al. 2002, Molnár-Láng et al. 2012) (see Figs. 13.1 and 13.2)

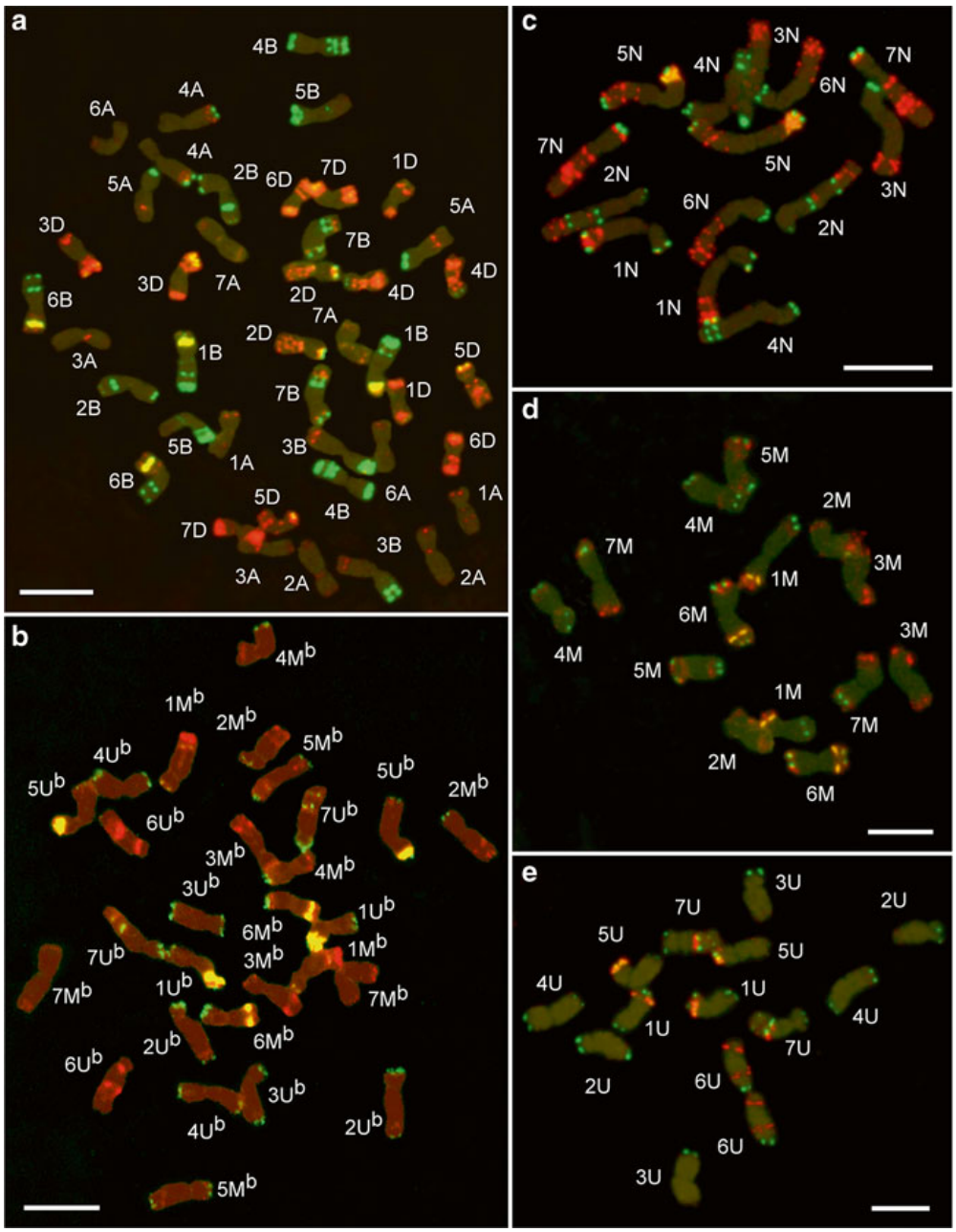


Fig. 13.1 Molecular cytogenetic identification of mitotic metaphase chromosomes in (**a**) *T. aestivum* cv. Chinese Spring ($2n=6x=42$; BBAADD); (**b**) *Ae. biuncialis* MvGB382 ($2n=4x=28$; $U^bU^bM^bM^b$); (**c**) *Ae. uniaristata* JIC2120001 ($2n=2x=14$; NN); (**d**) *Ae. comosa* MvGB1039 (2n = 2x = 14; MM); and (**e**) *Ae. umbellulata* AE740/03 ($2n=2x=14$; UU). Fluorescence *in situ* hybridization (FISH) was done using repetitive DNA probes for Afa family repeat (red), pSc119.2 repeat (green) and pTa71 repeat (yellow) and allowed identification of all chromosomes in the karyotypes. Scale bar = 10 μm

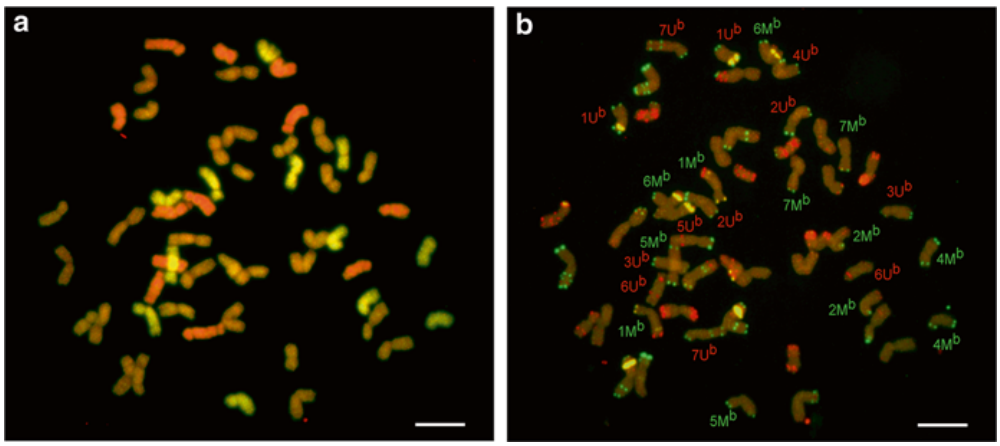


Fig. 13.2 Multicolor genomic *in situ* hybridization (mcGISH) using U- and M-genomic probes (**a**) and FISH with probes for DNA repeats (**b**) on mitotic metaphase chromosomes of a partial meristem root tip cell of wheat-*Ae. biuncialis* amphiploid plant. (**a**) mcGISH allows discrimination of U^b genome (red color), M^b genome (green color), and wheat (brown color) chromosomes. (**b**) FISH with probes for pSc119.2 repeat (green color), Afa family repeat (red color), and pTa71 repeat (yellow color) enables identification of all alien chromosomes in the wheat background. Scale bar = 10 μm

Characteristic FISH labeling patterns of HvT01 tandem repeat (Schubert et al. 1998), and the Triticeae-specific AT-rich tandem repeat pHvMWG2315 (Busch et al. 1995), permitted identification of all chromosomes in barley. In wheat genetic background, barley chromosomes could be discriminated with various combinations of repetitive DNA probes (Szakács and Molnár-Láng 2007). In rye, FISH with the 120-bp repeat family pSc119.2 together with pTa71 or AAC repeats identifies the whole chromosome complement (McIntyre et al. 1990; Szakács and Molnár-Láng 2008). In order to enrich chromosomes with diagnostic landmarks, microsatellite trinucleotide repeats (GAA, AAC, ACG) were found useful in wheat, barley, and rye (Cuadrado et al. 2008) as well as in *Aegilops* (Molnár et al. 2011a) and *Dasypyrum* (Grosso et al. 2012).

Inserts from DNA libraries cloned in a BAC (Bacterial Artificial Chromosome) vector were also tested to identify new repetitive sequences (both dispersed and tandem types), and to develop locus-specific cytogenetic markers (Zhang et al. 2004a). FISH with BAC clones (BAC FISH) was shown useful to discriminate the three subgenomes in hexaploid wheat (Zhang et al. 2004b), and for physical mapping of a powdery mildew-resistance gene (Yang et al. 2013). Unfortunately, BAC FISH suffers from the presence of dispersed repetitive DNA sequences in BAC clones, which often prevent localization of BAC clones to single loci. A possible solution is to use short single-copy probes free of repeats (Karafiátová et al. 2013).

Danilova et al. (2014) used wheat cDNAs as probes for FISH to develop cytogenetic markers specific for single-copy genic loci in wheat. They localized several cDNA markers on each of the 14 homoeologous chromosome arms and studied chromosome structure and homoeology in wild Triticeae species. The work revealed

1U-6U chromosome translocation in *Ae. umbellulata*, showed collinearity between the chromosome A of *Ae. caudata* and group-1 wheat chromosomes, and between chromosome arm 7S#3L of *Thinopyrum intermedium* and the long arm of the group-7 wheat chromosomes. A limitation inherent to performing FISH on condensed mitotic and meiotic chromosomes is the low spatial resolution. This can be improved by performing FISH on stretched mitotic chromosomes (Valárik et al. 2004), on extended DNA fibers (Fiber-FISH) (Jackson et al. 1998; Ersfeld 2004), and on hyper-expanded chromosomes obtained by flow cytometry (Endo et al. 2014).

Genomic *in situ* hybridization (GISH) uses genomic DNA as a probe (Schwarzacher et al. 1989) and permits determination of genomic constitution of allopolyploid Triticeae, and to detect alien chromatin introgressed into wheat. Combined with chromosome banding and/or FISH, the method allows location and identification of wheat–alien translocation breakpoints (Friebe et al. 1992, 1993; Jiang et al. 1993; Molnár-Láng et al. 2000, 2005; Liu et al. 2010; Kruppa et al. 2013). While cytogenetic methods are irreplaceable to verify genomic constitution in interspecific hybrids, the limited sensitivity and spatial resolution, and especially their laborious and time consuming nature seriously limit their suitability for large scale selection of wheat–alien introgressions. High-resolution and high-throughput methods are needed to increase the screening capacity and to identify micro-introgressions and chromosome breakpoints. These include the use of DNA markers and, more recently, DNA sequencing.

13.3.2 Molecular Markers

Morphological, isozyme, and seed storage protein markers were the first markers used in wheat–alien introgression breeding to identify and characterize alien chromosome addition lines (Guadagnuolo et al. 2001; Hart et al. 1980; Tang and Hart 1975). Because of their limited number, they were not suitable to reveal chromosomal rearrangements.

The restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPD) (Williams et al. 1990), and amplified fragment length polymorphism (AFLP) (Vos et al. 1995), were the first DNA markers used to characterize wheat–alien introgression lines (Fedak 1999), since they do not require prior sequence information. They were used in a number of studies to identify chromosome/chromosome-arm addition and substitution lines (Devos and Gale 1993; King et al. 1993; Hernández et al. 1996; Qi et al. 1996; Peil et al. 1998; Wang et al. 1995; Francki et al. 1997; Qi et al. 1997). Despite their temporal popularity, the markers suffered from some drawbacks. Their application was time-consuming, often labor-intensive and expensive, and they were not appropriate for high-throughput genotyping. Moreover, the low level of polymorphism revealed by RAPD markers, and low transferability/conversion of AFLP markers into STS markers, prevented the extensive use of these markers in wheat breeding (Gupta et al. 1999).

RFLPs became the molecular markers of choice for some time due to their codominance and locus specificity (Qi et al. 2007). Wheat RFLPs were used to develop high-resolution genetic and physical maps (Qi et al. 2004; Qi et al. 2003), characterize homoeology of alien chromosomes, and reveal their rearrangements relative to wheat (Devos et al. 1993; Devos and Gale 1993; Zhang et al. 1998; McArthur et al. 2012). RFLP markers identified cryptic alien introgressions where cytogenetic techniques failed (Yingshan et al. 2004), such as the T5DL.5DS-5M^{g}S wheat-*Ae. geniculata* translocation conferring resistance to leaf rust and stripe rust (Kuraparthy et al. 2007). With the advances in molecular biology, informative but cumbersome to use RFLP markers were converted to PCR-based markers such as the sequence-tagged site (STS) markers, which were more suitable for tagging interesting genes (Cenci et al. 1999; Seyfarth et al. 1999; Langridge et al. 2001).

Transposable elements, randomly distributed in nuclear genomes have also been used as molecular markers (Queen et al. 2003; Nagy and Lelley 2003). The sequence-specific amplified polymorphism (S-SAP) technology (Waugh et al. 1997) amplifies regions representing flanking genomic sequences of individual retrotransposons. The advantages of S-SAP for studying genetic diversity are higher amount of accessible polymorphism (Waugh et al. 1997), the markers are more evenly distributed throughout the genome (Nagy and Lelley 2003), and the estimated genetic distances are more consistent with physical mapping (Ellis et al. 1998). Nagy et al. (2006) used the short interspersed nuclear element (SINE) Au identified in *Ae. umbellulata* (Yasui et al. 2001) to develop S-SAP markers specific for U- and M-genome chromosomes of *Aegilops* (Nagy et al. 2006).

Simple Sequence Repeat (SSR) markers (Tautz 1989), or microsatellite markers, were the next generation of molecular markers employed in wheat–alien introgression breeding (Mohan et al. 2007; Bandopadhyay et al. 2004; Yu et al. 2004; Gupta et al. 2003). Efficient development of SSRs requires genomic sequence information, and thus they were developed concomitantly with expressed sequence tags (ESTs), cDNA and BAC libraries. A list of genomic resources currently available for Triticeae is given in Table 13.1.

Together with cDNA libraries and draft genome sequences of barley, bread wheat, *Ae. tauschii* and *T. urartu* (Table 13.2), ESTs are currently the most abundant type of sequence information available for not less than 25 species from 15 Triticeae genera. The release of 16,000 EST loci mapped to chromosome deletion bins (Qi et al. 2004) provided excellent resource for development of markers from specific chromosome regions and helped designing locus-specific markers. Because of the genic and thus conserved nature of ESTs, EST-derived SSR markers are transferable between Triticeae species (Gupta et al. 2008). As ESTs and cDNA resources are much less abundant in other Triticeae, e.g., *Elymus*, *Aegilops* and *Leymus*, numerous studies profited from the high transferability of wheat EST-derived SSR markers across distantly related species for comparative mapping, trait-marker associations and to carry out evolutionary studies to establish the phylogenetic relationships among the wild relatives of wheat and between them and bread wheat (Adonina et al. 2005; Jing et al. 2007; Kroupin et al. 2012).

The conserved orthologous set (COS) markers allowed identification of orthologous regions between wild species and wheat in order to facilitate alien gene-transfer

Table 13.1 Genomic resources available for Triticeae species

Genus (no. of taxonomy entries in NCBI)	Bio Project[a]	Number of genes	Number of ESTs	BAC libraries	cDNA clones	Probe[b]	Map data[c]	SRA[d]	GSS[e]	Genome[f]
Aegilops (42)	35	1172	4546	8	2303	787	4	161	5172	1
Agropyron (16)	0	4	17	–	–	0	–	1	0	1
Amblyopyrum (3)	–	–	–	–	–	–	–	–	–	–
Anthosachne (10)	–	–	–	–	–	–	–	–	–	–
Australopyrum (6)	–	–	–	–	–	–	–	–	–	–
Avena (35)	11	28	79,657	–	–	11,542	24	–	3063	–
Campeiostachys (11)	–	–	–	–	–	–	–	–	–	–
Connorochloa (2)	–	–	–	–	–	–	–	–	–	–
Critesion (4)	–	–	–	–	–	–	–	–	–	–
Crithopsis (2)	–	–	–	–	–	–	–	–	–	–
Dasypyrum (3)	–	–	–	–	–	–	–	–	14	–
Douglasdeweya (3)	–	–	–	–	–	–	–	–	–	–
Elymus (116)	1	–	45,580	–	–	–	–	–	–	1
Eremopyrum (5)	-	–	–	–	–	–	–	–	–	–
Festucopsis (3)	–	–	–	–	–	–	–	–	–	–
Haynaldia (2)	–	–	10	–	–	–	–	–	13	–
Henrardia (5)	–	–	–	–	–	–	–	–	–	–
Heteranthelium (4)	–	–	–	–	–	–	–	–	–	–
Hordelymus (2)	–	–	–	–	–	–	–	–	–	–
Hordeum (62)	148	717	840,120	2	89,452	11,196	76	1894	574,028	4
Hystrix (5)	–	147	–	–	–	–	–	50	–	–
Kengyilia (22)	–	–	–	–	–	–	–	2	–	–
Leymus (50)	4	–	30,749	–	–	1853	3	6	13	–
Lophopyrum (5)	2	–	2	–	–	56	–	1	–	–

(continued)

Table 13.1 (continued)

Genus (no. of taxonomy entries in NCBI)	Bio Project[a]	Number of genes	Number of ESTs	BAC libraries	cDNA clones	Probe[b]	Map data[c]	SRA[d]	GSS[e]	Genome[f]
Pascopyrum (2)	–	–	–	–	–	–	–	–	–	–
Peridictyon (2)	–	–	1	–	–	–	–	–	–	–
Psammopyrum (2)	–	–	–	–	–	–	–	–	–	–
Psathyrostachys (16)	1	–	–	–	–	–	–	1	44	–
Pseudoroegneria (9)	–	–	–	–	–	–	–	–	–	–
Secale (16)	21	113	15,903	2	6617	1091	12	36	2956	–
Stenostachys (4)	–	–	–	–	–	–	–	–	–	–
Taeniatherum (6)	–	–	2	–	–	–	–	–	–	–
Thinopyrum (12)	4	–	2385	–	–	–	–	3	7	–
Triticum (84)	239	3170	1,358,421	16	10,527	21,164	69	2558	72,374	4
× *Aegilotriticum* (14)	1	–	–	–	–	–	–	–	–	–
× *Triticosecale* (10)	3	–	11	–	–	–	–	–	8	–
× *Tritordeum* (6)	–	–	4	–	–	57	–	–	11	–

The information in this table was collected from NCBI taxonomy (http://www.ncbi.nlm.nih.gov/taxonomy) and GrainGene (http://wheat.pw.usda.gov/GG3/) databases in May 2015. Triticeae genera comprising cultivated species are underlined

[a]Projects initiated in the fields of genomics, functional genomics and genetic studies (NCBI)

[b]Public registry of nucleic acid reagents designed for use in a wide variety of biomedical research applications (NCBI)

[c]Genetic and physical maps available for Triticeae (GrainGene database)

[d]Sequence Read Archive (NCBI) stores sequencing data

[e]Genome Survey Sequences (NCBI) is a collection of unannotated short single-read primarily genomic sequences from GenBank including random survey sequences, clone-end sequences and exon-trapped sequences

[f]Genome (NCBI) reference whole genomes sequencing information, both completely sequenced organisms and those for which sequencing is in progress

Table 13.2 Whole genome sequencing projects in cereals

Species/cultivar	Genome size (1C)	Sequence description	Consortium/team
Oryza sativa ssp. *japonica*	400–430 Mbp	Pseudomolecule	International Rice Genome Sequencing Project (2005)
Zea maize cv. B73	2.4 Gbp	Pseudomolecule	Schnable et al. (2009)
Sorghum bicolor cv. Moench	750 Mbp	Whole-genome draft assembly	Paterson et al. (2009)
Brachypodium distachyon inbread line Bd21	~355 Mbp	Pseudomolecule	The International Brachypodium Initiative (2010)
Hordeum vulgare cv. Morex	~5.3 Gbp	Whole-genome draft assembly	The International Barley Genome Sequencing Consortium (2012)
Aegilops tauschii ssp. *strangulata* accession AL8/78	4.02 Gbp	Whole-genome draft assembly	Luo et al. (2013)
Triticum urartu accession G1812	4.94 Gbp	Whole-genome draft assembly	Ling et al. (2013)
Triticum aestivum cv. Chinese spring (CS) 3B chromosome of *Triticum aestivum* cv. CS	~16 Gbp ~16 Gbp (~1 Gbp)	5× whole-genome draft assembly Chromosome-based draft genome assembly Reference sequence assembly of chromosome 3B	Brenchley et al. (2012) IWGSC (2014) Choulet et al. (2014)

through a better characterization of the potentially recombining regions (Molnár et al. 2013). As the COS markers are PCR based and span exon–intron junctions, they are conserved enough to be transferrable across genera, while the intron sequences provide relatively high polymorphism that allows variants of genes to be discriminated (e.g., between species). Although these markers present interesting tools to support alien-wheat gene transfer, they remain underexploited in this area.

13.3.3 High-Throughput Genotyping

Diversity Arrays Technology (DArT) markers were initially developed as micro-array hybridization-based sequence-independent marker system, and allowed screening thousands of polymorphic loci in a single assay at low cost per data point (Jaccoud et al. 2001). Among other, DArT markers were used to develop high-density genetic map of wheat × wild emmer (Peleg et al. 2008). A new version of DArT marker technology (DArT-seq) is based on next-generation sequencing where the polymorphisms are genotyped by sequencing. Because of its advantages,

DArT has been employed extensively in genetic mapping, genotyping, and diversity assessment in wheat (Cabral et al. 2014; Jighly et al. 2015; Bentley et al. 2014; Yu et al. 2014; Colasuonno et al. 2013; Iehisa et al. 2014), and more recently in its wild and cultivated relatives (Montilla-Bascón et al. 2015; Kalih et al. 2015; Castillo et al. 2014; Bolibok-Brągoszewska et al. 2014; Alheit et al. 2014; Yabe et al. 2014; Cabral et al. 2014; Jing et al. 2009).

The advent of the next generation sequencing technologies changed the paradigm of wheat genetics and genomic and led to the development of Single Nucleotide Polymorphism (SNP) markers. Various platforms have been developed for wheat genotyping such as the 9K and 90K Illumina iSelect® platforms with 9000 and 90,000 SNP markers, respectively (Cavanagh et al. 2013; Wang et al. 2014), the Illumina Infinium® platform (up to 1,000,000 SNP markers), as well as the Axiom® 820K and 35K arrays (with up to 820,000 and 35,000 features) (http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/axiom_download.php). These platforms provide tools to obtain detailed information on germplasm diversity and characterize allelic variation. However, low representation of wild wheat relatives in the SNP design process may limit the utility of the platforms in wheat alien introgression breeding (Wulff and Moscou 2014). Consequently, a few studies made use of SNP molecular markers to support alien gene transfer in wheat (Tiwari et al. 2014) and very few SNPs derived from wild species are available.

Due to the low cost per data point and ease of development, Kompetitive Allele Specific PCR (KASP) SNP markers (He et al. 2014), a genotyping technology based on allele-specific oligo extension and fluorescence resonance energy transfer for signal generation, are becoming popular and are used in large-scale projects (Petersen et al. 2015). KASP markers can genotype SNP polymorphism, deletions and insertions variations, and have been used in screening wheat–alien hybrids and their back-crossed derivatives to detect recombinants and isolate desired introgressions (King et al. 2013). In order to promote the use of KASP markers, it is important to generate new genomic sequences from wild relatives of wheat.

13.3.4 Genome Sequencing

13.3.4.1 Whole Genome Approaches

Despite the importance of Triticeae species for the humankind (Feuillet et al. 2008), attempts to sequence their genomes were delayed due to the size and complexity. The nuclear genome of bread wheat comprises three structurally similar (homoelogous) subgenomes A, B, and D, and with the size of about 17 Gb/1C, it is 40 times bigger than rice (0.43 Gb) and 126 times bigger than *Arabidopsis thaliana* (0.135 Gb). As the other Triticeae genomes, it is highly redundant and composed mostly from various classes of repetitive DNA sequences (IWGSC 2014).

High throughput of the next generation sequencing technologies makes it possible to sequence even the biggest genomes. However, the problem is to assemble and

order the short reads thus obtained (IWGSC 2014). Due to large genome complexity and sequence redundancy, high-quality reference genome assemblies are not yet available for any of the Triticeae species. To date, only draft genome sequences are available for barley (The International Barley Genome Sequencing Consortium 2012), *T. urartu* (Ling et al. 2013)—a progenitor of the A genome of bread wheat, *Ae. tauschii* (Luo et al. 2013)—a D genome progenitor of bread wheat, as well as the whole genome shotgun assembly of hexaploid bread wheat (Brenchley et al. 2012) (see Table 13.2).

Due to their nature, draft sequence assemblies are only partial representations of the genomes, often accounting for less than 50 % of their estimated size. A significant part of expressed genes may be absent, which may compromise efforts with gene discovery and cloning, while the fragmentation of genome sequence and large numbers of unanchored contigs hamper comparative genome analyses.

Despite their preliminary nature, draft genome sequences provided useful insights into the Triticeae genome organization, evolution, and function. They were useful to develop protein-coding gene models, analyze genome organization, assess recombination rates along chromosomes, and characterize synteny and collinearity with other species (Ling et al. 2013; Luo et al. 2013; The International Barley Genome Sequencing Consortium 2012). They served as templates to characterize agronomically important genes and develop genome-specific molecular markers for plant breeding (Ling et al. 2013). The utility and extensive use of whole genome sequences from the main Triticeae crops confirm the need for such resources in wild wheat relatives. Although it may not be possible to sequence genomes of all wild species employed in wheat alien introgression breeding, efforts should be made to obtain as much information on their genomes as possible in order to understand better the genome relationships among Triticeae.

13.3.4.2 Reduced-Complexity Sequencing

One approach to facilitate sequencing and assembly of the huge Triticeae genomes is to reduce sample complexity prior to sequencing. Various strategies have been developed to achieve this, and can be classified into two groups: (1) Transcriptome sequencing and sequence capture approaches, which sequence only certain parts of genomes, and (2) the chromosome-centric approaches, which reduce the complexity in a lossless way by dissecting genomes to small parts (chromosomes and chromosome arms) that are sequenced and assembled separately.

Sequencing conserved genic portions of genomes enables development of cross-species transferable tools, and facilitates functional understanding of important traits. Haseneyer et al. (2011) sequenced transcriptome in five winter rye inbred-lines and identified over 5000 SNPs between the transcriptomes that were subsequently used for genotyping 54 inbred lines using SNP genotyping array. This analysis does not require prior knowledge of genome sequence and allows large-scale molecular marker development for high-throughput genotyping. A recent analysis of *Agropyron cristatum* transcriptome permitted identification of 6172

unigenes specific to *A. cristatum*, including many stress-resistant genes and alleles potentially useful in wheat improvement (Zhang et al. 2015).

Another option to reduce sequencing efforts are sequence-capture approaches, which are used to enrich samples for sequences of interest before carrying out NGS. They are based on hybridization of target sequences to bait probes in solution, or on solid support. This approach usually necessitates preliminary sequence information. However, since it allows high level of mismatches, it permits capturing diverged sequences. Known sequences from more characterized species such as wheat, barley, *Brachypodium*, and rice can be employed to discover uncharacterized sequences from related species and varieties. Accordingly, Jupe et al. (2013) developed an exome capture for nucleotide-binding leucine-rich repeat (NB-LRR) domain for the so-called Resistance gene enrichment Sequencing (RenSeq) in potato. Their work resulted in discovery of 317 previously unannotated NB-LRRs and the method could aid in discovery of new resistance genes in wild relatives of wheat (Wulff and Moscou 2014).

Alternative approach to reduce complexity of large and polyploid genomes is to isolate chromosomes by flow cytometric sorting and sequence them individually (Fig. 13.3). This strategy is called chromosome genomics (Doležel et al. 2007, 2014) and has been adopted by the IWGSC for the bread wheat genome sequencing (IWGSC 2014). The method, originally developed in *Vicia faba* (Doležel et al. 1992), relies on cell cycle synchronization of meristem root tip cells of young seedlings and their accumulation at mitotic metaphase. After mild formaldehyde fixation, intact chromosomes are released into a buffer solution by mechanical homogenization of root tips. Chromosome samples are stained by a DNA fluorochrome DAPI and classified at rates of several thousand per second according to their relative DNA content using flow cytometry. Chromosomes that differ in DNA content from other chromosomes form distinct peaks on histograms of DNA content (flow karyotypes). Such chromosomes, can be sorted individually at rates of about 20 s^{-1}, and several hundred thousand chromosomes of the same type can be collected in one day.

In a majority of species, chromosomes have similar DNA content and cannot be discriminated after DAPI staining alone. The most frequent approach to overcome this difficulty has been the use of cytogenetic stocks in which the size of one or more chromosomes has been changed so that the chromosome of interest can be discriminated and sorted. The stocks included chromosome translocations (Kubaláková et al. 2002), deletions (Kubaláková et al. 2005), alien chromosome addition (Kubaláková et al. 2003) and alien chromosome arm additions (Suchánková et al. 2006). As such stocks are not available for many species, it is important that Giorgi et al. (2013) developed a protocol termed FISHIS, to fluorescently label repetitive DNA on chromosomes prior to flow cytometric analysis. This approach permits discrimination of chromosomes, which have the same or very similar relative DNA content (Fig. 13.3), and has been used successfully to sort chromosomes in *Ae. umbellulata*, *Ae. comosa*, *Ae. speltoides*, and *Ae. markgrafii* (in preparation).

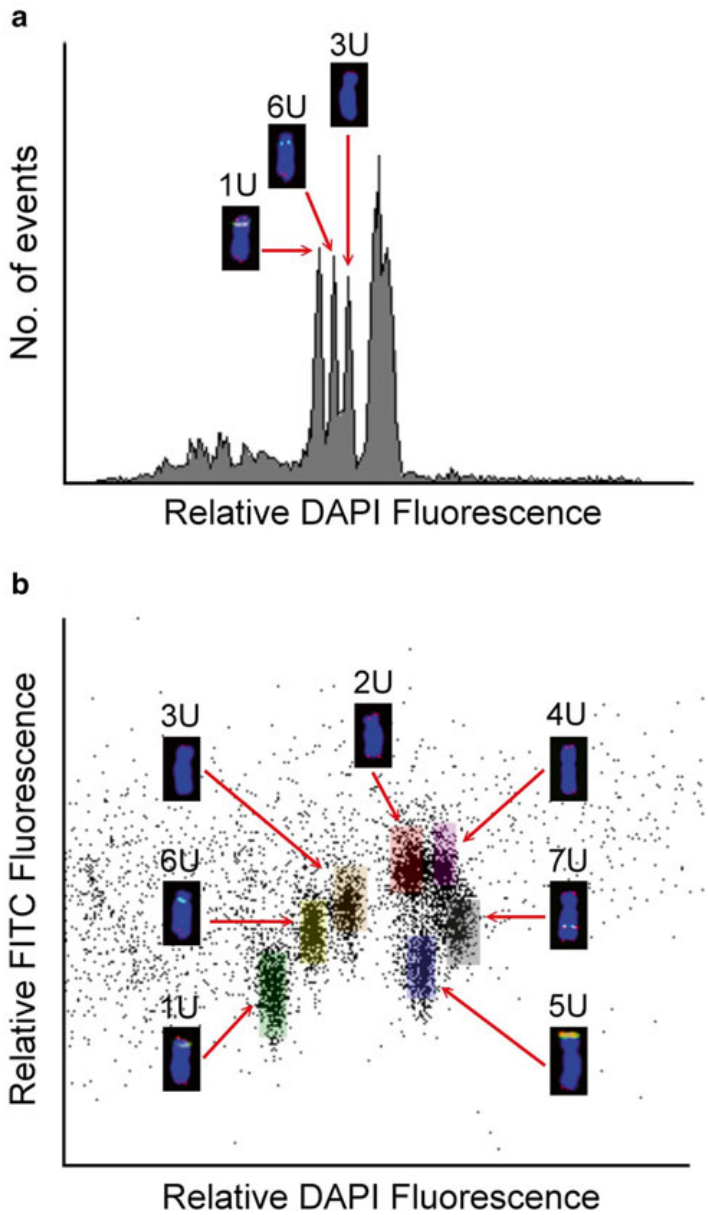


Fig. 13.3 Mono- (**a**) and biparametric (**b**) flow cytometric analysis and sorting of mitotic metaphase chromosomes from *Ae. umbellulata* ($2n=2x=14$; UU). (**a**) Monoparametric analysis of chromosomes stained by DAPI results in a histogram of relative fluorescence intensity (flow karyotype) in which three peaks representing chromosomes 1U, 6U and 3U are discriminated. The remaining four chromosomes form a composite peak and cannot be sorted individually. Biparametric analysis of chromosomes stained by DAPI and with GAA repeats labeled by FITC results in a bivariate flow karyotype on which all seven chromosomes (colored regions) can be discriminated and flow-sorted at a purity of 90–99 %

Table 13.3 List of Triticeae species in which flow cytometric chromosome sorting has been reported (adapted from Doležel et al. (2014))

Genus	Species	Common name	*n*	Reference[a]
Aegilops	*biuncialis*	Goatgrass	14	Molnár et al. (2011b)
	comosa		7	Molnár et al. (2011b)
	cylindrica		14	Molnár et al. (2015)
	geniculata		14	Molnár et al. (2011b); Tiwari et al. (2014)
	markgrafii		7	Molnár et al. (2015)
	speltoides		14	Molnár et al. (2014)
	triuncialis		14	Molnár et al. (2015)
	umbellulata		7	Molnár et al. (2011b)
Avena	*sativa*	Oat	21	Li et al. (2001)
Dasypyrum	*villosum*	Mosquito Grass	7	Grosso et al. (2012); Giorgi et al. (2013)
Hordeum	*vulgare*	Barley	7	Lysák et al. (1999); Lee et al. (2000); Suchánková et al. (2006); **Mayer et al. (2009, 2011)**
Secale	*cereale*	Rye	7	Kubaláková et al. (2003); **Bartoš et al.** (2008); **Martis et al.** (2013)
Triticum	*aestivum*	Bread wheat	21	Wang et al. (1992); Schwarzacher et al. (1997); Lee et al. (1997); Gill et al. (1999); Vrána et al. (2000); Kubaláková et al. (2002); Giorgi et al. (2013); **Hernandez et al.** (2012); **IWGSC** (2014); **Helguera et al.** (2015); Tanaka et al. (2014); Sergeeva et al. (2014); **Lucas et al.** (2014); **Berkman et al.** (2011)
	durum	Durum wheat	14	Kubaláková et al. (2005); Giorgi et al. (2013)
	urartu		7	Molnár et al. (2014)

[a]Reports on chromosome sequencing are printed in bold

To date, chromosome flow-sorting has been reported in at least 29 plant species, including 15 Triticeae (Doležel et al. 2014; Table 13.3). High purity in the sorted fractions and high molecular weight DNA of flow-sorted chromosomes makes them ideal substrate for downstream applications such as PCR-based analysis, development of markers, BAC-vector cloning and construction of optical maps (for review see (Doležel et al. 2014)). Chromosomal DNA can be sequenced or used for other applications either directly, if a sufficient number of chromosomes is sorted, or after representative amplification (Šimková et al. 2008). It is now even possible to sequence a single flow-sorted chromosome (Petr Cápal pers. comm.). The latter is particularly important in cases when the chromosome of interest cannot be discriminated from other chromosomes in karyotype, or if the focus is on the analysis of structural chromosome heterozygosity and allele phasing.

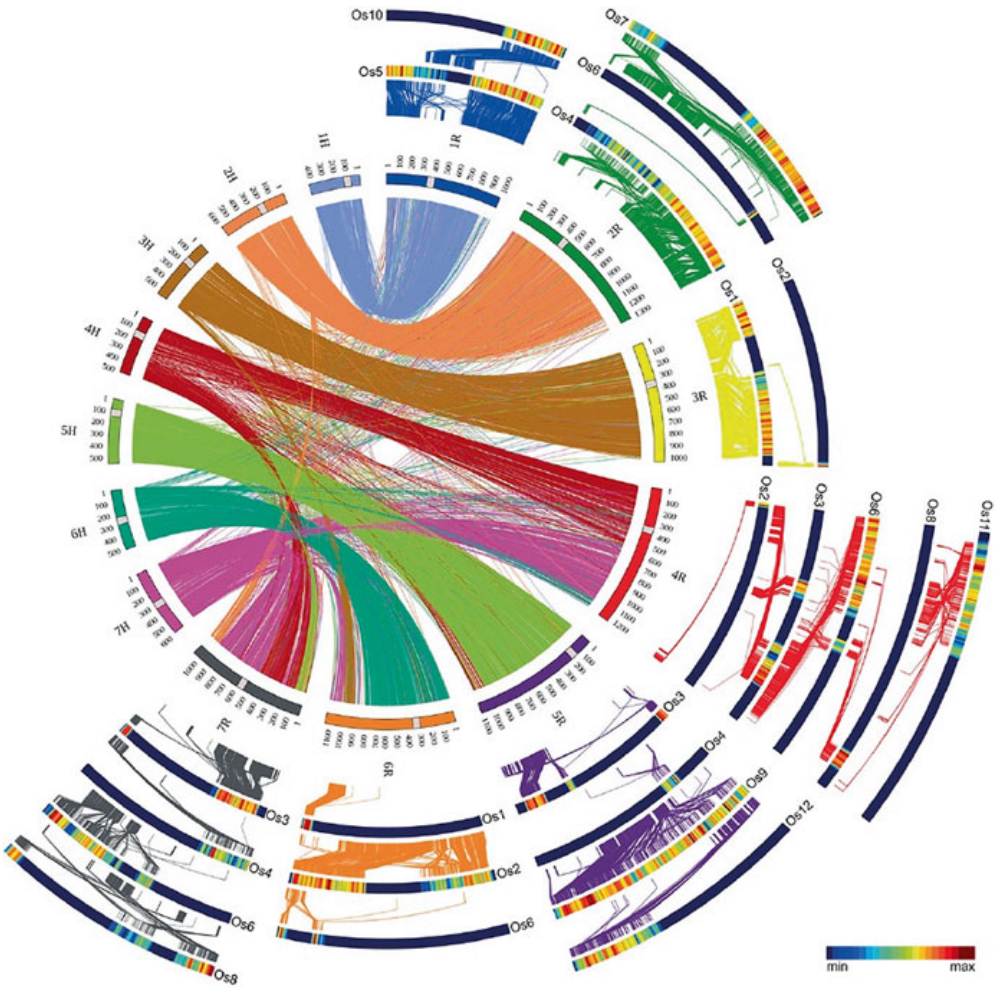


Fig. 13.4 Next-generation sequencing of flow-sorted rye chromosomes allowed characterization of synteny between rye, barley, and rice genomes. Collinearity of the rye and barley genomes is depicted by the inner circle of the diagram. Rye (1R–7R) and barley (1H–7H) chromosomes were scaled according to the rye genetic and barley physical map, respectively. Lines (colored according to barley chromosomes) within the inner circle connect putatively orthologous rye and barley genes. The outer partial circles of heat map colored bars illustrate the density of rice genes hit by chromosome sequencing reads of the corresponding rye chromosomes. Conserved syntenic blocks are highlighted by yellow-red-colored regions of the heat maps. Putatively orthologous genes between rye and rice are connected with lines (colored according to rye chromosomes) and centromere positions are highlighted by grey rectangles. Martis et al., Plant Cell 25: 3685–3698, 2013. www.plantcell.org Copyright American Society of Plant Biologists. Reproduced with permission

For example, BAC-end sequences obtained using 1RS-specific BAC library were used to develop Insertion Site-Based Polymorphism markers (ISBP) specific for 1RS and to identify loci carrying microsatellites suitable for the development of 1RS-specific SSR markers (Bartoš et al. 2008). Next-generation sequencing flow-sorted chromosomes of rye enabled establishing linear gene order model com-

prising over 22 thousand genes, i.e. 72 % of the detected set of 31,000 rye genes. Chromosome sequencing together with transcript mapping and integration of conserved synteny information of Brachypodium, rice and sorghum enabled a genome-wide high-density comparative analysis of grass genome synteny (Fig. 13.4).

The chromosome genomics approach has been particularly fruitful in genomics of wheat. The chromosome-based draft sequence of bread wheat was obtained by sequencing flow-sorted chromosome arms (except of chromosome 3B), each of them representing only 1.3–3.3 % of the genome. Chromosome arms were sequenced with Illumina technology and the reads were assembled to contigs representing 10.2 Gb (61 %) of the genome with a L50 of repeat-masked assemblies ranging from 1.7 to 8.9 kb. A total of 133,090 loci homologous to related grass genes were classified as high-confidence gene calls. Out of them, 93.3 % were annotated on individual chromosome arm sequences, and 53.2 % were located on syntenic chromosomes compared to brachypodium, rice and sorghum. In total, 81 % raw reads and 76.6 % assembled sequences contained repeats, explaining the difficulty of assembling such genomes from short sequence reads. As demonstrated in chickpea, chromosome genomics can be coupled with whole genome next-generation sequencing to validate whole genome assemblies (Ruperao et al. 2014). This powerful combination could speed up production of good quality whole genome assemblies in wild wheat relatives.

Chromosome genomics was also shown useful to characterize chromosome segments of alien origin, develop markers from these regions, and support cloning alien genes of interest. In a pioneering study, Tiwari et al. (2014) sequenced DNA from flow-sorted short arm of chromosome $5M^g$ of *Ae. geniculata* to develop genome-specific SNP markers. The markers allowed development of two SNP markers identifying introgression of a segment of $5M^g$ to wheat chromosome 5D carrying resistance to leaf rust (*Lr57*) and stripe rust (*Yr40*) (Fig. 13.5). In order to simplify the identification of alien chromatin introgressed into wheat, Abrouk (pers. comm.) developed a method based on comparative analysis. Briefly, using the linear gene order map of a recipient wheat chromosome (IWGSC 2014) and the sequence of flow-sorted chromosome carrying alien introgression, the density of orthologs is calculated along the wheat chromosome. The variation in density makes it possible to detect the alien segment. This approach has been validated recently in wheat *T. aestivum* cv. *Tahti—T. militinae* introgression line 8.1 (Jakobson et al. 2006, 2012), which carries a major QTL for powdery mildew resistance on the distal part of the long arm of chromosome 4A (Michael Abrouk pers. comm.)

13.4 Functional Aspects of Alien Gene Transfer

When introducing alien genes to wheat, the function of introgressed chromosomes or chromosome segments and their interaction with the host genome needs to be considered. It may occur at different levels and concern chromosome behavior during meiosis, changes in chromosomes structure and genome organization, as well as gene expression. Understanding the interaction between the host and alien genomes,

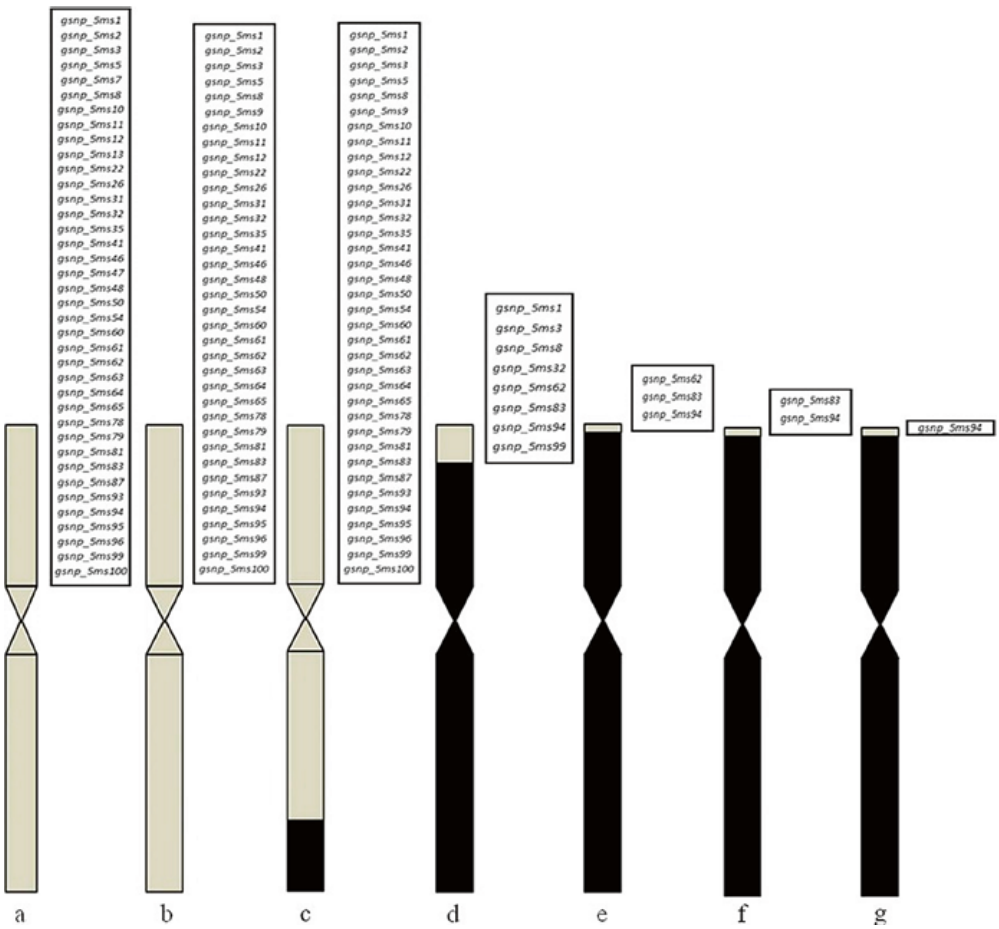


Fig. 13.5 Distribution of validated 5M^{g}S-specific SNPs developed from flow-sorted ditelosomic 5M^g in different alien introgression-based addition, translocation, and released wheat lines. (**a**) disomic addition line TA7657, (**b**) disomic substitution line TA6675, (**c**) translocation line TA5599, (**d**) terminal translocation line TA5602, (**e**) TA5602 (with very small 5M^g segment), (**f**) SNPs validated in germplasm KS11WGGRC53-J and (**g**) SNP validated in germplasm KS11WGGRC53-O. Tiwari et al., BMC Genomics 15: 273, 2014. http://www.biomedcentral.com/bmcgenomics BioMed Central Ltd. Reproduced with permission

the evolution of this relationship from the moment of F1 hybrid formation to a stabilized wheat–alien introgression line, and the way the final equilibrium impacts the performance of the introgression line may contribute to the success of alien gene transfer in wheat improvement.

13.4.1 Interaction Between Host and Donor Genomes

Alien gene transfer involves hybridization and creation of interspecific hybrids, followed by genome duplication to establish fertile amphiploids. A consequence is a shock for both genomes, which may result in activation of mobile genetic elements,

various structural changes and lead to changes in epigenetic status of chromatin and novel patterns of gene expression (Comai 2000).

Elimination of specific sequences is commonly reported as rapid genomic rearrangement accompanying allopolyploidization in wheat. The changes include elimination of noncoding and low-copy DNA sequences, and gain of novel fragments (Feldman et al. 1997; Liu et al. 1998). Elimination of rye-specific fragments often representing transposable elements (TEs) and their derivatives was observed in allopolyploid triticales (Ma and Gustafson 2006, 2008; Bento et al. 2008). The analysis of a newly synthesized triticale (Bento et al. 2008; Han et al. 2003) revealed rapid changes in coding sequences upon the induction of allopolyploidy, but the changes did not extend to alterations discernible at cytological level. The molecular mechanisms underlying genome reorganization are not yet fully understood (Tayalé and Parisod 2013). 'Genomic stress' due to polyploidization may activate TEs and promote their proliferation and mobility. At the same time, massive elimination in a TE family-specific manner may be observed (Comai et al. 2003; Parisod and Senerchia 2012). It seems that the degree of TE sequence divergence between progenitors correlates with the degree of restructuring in polyploid TE fractions (Senerchia et al. 2014).

A general observation made in newly created polyploids and synthetic allotetraploids, including wheat, is a change in gene expression immediately after polyploidization (Kashkush et al. 2002; Levy and Feldman 2004). Both genetic and epigenetic mechanisms may alter gene expression (Lynch and Conery 2000; Lee and Chen 2001; Osborn et al. 2003; Soltis et al. 2004). The analysis of cytosine methylation in *Aegilops–Triticum* F1 hybrids and their derivative allotetraploids revealed 13 % of the loci with altered patterns of methylation affecting both repetitive DNA and low-copy DNA (Xiong et al. 1999; Shaked et al. 2001). In leaves of *Arabidopsis* autopolyploids and allotetraploids and their progenitors, Ng et al. (2012) could associate rapid changes in gene expression with quantitative proteomic changes, suggesting rapid changes in post-transcriptional regulation and translational modifications of proteins as a consequence of polyploidization.

Epigenomic rearrangements after allopolyploidization seem to be involved in the processes of uniparental chromosome elimination, a phenomenon observed frequently in interspecific hybrids between *T. aestivum* and *H. bulbosum* (Bennett et al. 1976), *H. vulgare* (Islam et al. 1981) and *Zea mays* (Laurie and Bennett 1986). The loss of centromere-specific histone H3 (CENH3) caused centromere inactivation and triggered mitosis-dependent uniparental chromosome elimination in unstable *H. vulgare*×*H. bulbosum* hybrids (Sanei et al. 2011). Bento et al. (2010), found that chromosome structural rearrangements were more drastic in wheat–rye disomic addition lines than in triticale, indicating that the lesser the amount of rye genome introgressed into wheat, the higher the likelihood of wheat chromosome breakage, chromosome elimination, and chromosome structural rearrangement, including sequence-specific elimination, translocations and TE movement (Fu et al. 2013).

13.4.2 Alien Gene Expression

Various studies indicate complex relationships between the alien and host genes (Pumphrey et al. 2009; Jeffrey Chen and Ni 2006; Bougas et al. 2013; Wu et al. 2015; Yoo et al. 2013; Wulff and Moscou 2014) and, as a result, in some cases alien genes may not function as expected. For example, weaker effect in the wheat background as compared to the wild species was observed in studies involving resistance gene transfer (Wulff and Moscou 2014; Chen et al. 2005; Riley and Chapman 1958; Riley and Macer 1966). One explanation may be that the introgressed genes are involved in polygenic resistance together with other loci, which are not introgressed simultaneously. However, in some cases, resistance genes had no effect at all, as was the case of resistance to wheat leaf rust (*Puccinia triticina* Erikss.) introduced to wheat from rye (Riley and Macer 1966). It seems that the polyploid status of wheat itself may impact alien gene expression. When Kerber and Dyck (1973) transferred stem rust resistance from diploid einkorn wheat (*T. monococcum* L.) to tetraploid durum and hexaploid bread wheat, a progressive loss of the resistance with increasing ploidy from diploid to hexaploid was observed. Chen et al. (2005) described different levels of scab resistance in progenies that involved the same wheat-*Leymus racemosus* alien chromosome translocation, or the same alien chromosome addition, possibly related to other components of resistance in the genetic background.

Suppression of resistance due to negative interaction of homoeologous and non-homoeologous loci between genomes is another effect observed in hexaploid wheat, and the examples include a conserved gene on chromosome arm 7DL that suppresses stem rust resistance, and suppression of powdery mildew locus *Pm8* by *Pm3* locus (Kerber and Aung 1999; Wulff and Moscou 2014). The suppression of introgressed *Pm8* resistance gene by its *Pm3* host ortholog in some wheat–rye 1BL.1RS translocation lines was not due to gene loss, mutation or gene silencing (Hurni et al. 2014). A coexpression analysis of *Pm8* and *Pm3* genes in *Nicotiana benthamiana* leaves followed by co-immunoprecipitation analysis showed that the two proteins interact and form a heteromeric complex, which might result in inefficient or absent signal transmission for the defense reaction. Stirnweis et al. (2014) suggested that the frequently observed failure of resistance genes introduced from the secondary gene pool into polyploid crops could be the result of the expression of closely related NB-LRR-resistance genes or alleles in the host genome, leading to dominant-negative interactions through a posttranslational mechanism involving LRR domains. A recent study showed that genes with low similarity between rye sequences and their closest matches in the *Triticum* genome have a higher probability to be repressed or deleted in the allopolyploid genome (Khalil et al. 2015).

13.4.3 Spatial Genome Organization and Function

Little is known how alien chromosome(s) and/or translocated alien chromosome segments influence behavior and position of wheat chromosomes within the 3D space of interphase nucleus, how the position and behavior of alien chromosome differs from that in the nucleus of donor wild relative, and how changes in chromosome position influence gene expression of wheat and alien genes. Numerous studies in human and mouse indicate that chromosome territories are not randomly positioned in the nucleus (Gibcus and Dekker 2013). Small and gene-rich chromosomes localize near the center of nucleus, whereas larger and less-gene-rich chromosomes are more frequently located near the nuclear periphery. In plants, however, 3D-nuclear genome organization has been studied only in a few cases and mostly in *Arabidopsis* (Schubert et al. 2014; Grob et al. 2014) and rice (Mukhopadhyay et al. 2013) with small genomes, whose interphase organization may differ from that of large genomes. The results obtained in rice (Mukhopadhyay et al. 2013) correlated transcriptional regulation with alteration in nucleosome positioning, histone modifications and gene looping, but not DNA methylation. A recent observation using 3D-FISH in wheat–rye chromosome arm introgression lines indicated that the rye alien chromosomes were positioned at the periphery of nuclei (Veronika Burešová, pers. comm.). These preliminary results are consistent with the general observation of negative regulation of the expression of the alien genes introgressed in wheat.

13.5 Concluding Remarks

During more than one century of wheat–alien introgression breeding, a significant progress has been made in developing strategies to produce hybrids of wheat with distant relatives, in devising chromosome engineering techniques to integrate alien-chromosome segments into wheat genome, in the improvement of cytogenetic techniques to identify and characterize introgressed chromatin, and in phenotypical characterization of new introgression lines. These advances led to development of a formidable panel of introgression lines of various types and from a number of wild wheat relatives, carrying important traits. Nevertheless, only a small number of commercially successful wheat cultivars benefitted from these advances, and the potential of alien introgression breeding remains underused.

In order to fully explore it and benefit from the extant genetic diversity of wild wheat relatives, implementation of improved and novel approaches and tools is needed. It is fortunate that new methods of cytogenetics, genomics and phenomics are becoming available for better and, in case of genomics and phenomics, high-throughput characterization of genetic diversity, and identification of donors of important traits. On the other hand, improvement of chromosome engineering methods and better knowledge of molecular mechanisms controlling meiotic recom-

bination are needed to facilitate successful introgression of alien chromatin. This will require a better knowledge of genome structure of wild relatives to assess chances for chromosome recombination and predict its outcomes, in order to decide the best experimental approach to be applied.

The advances in DNA sequencing and DNA marker technologies make it possible to compare genomic organization of wheat and wild relatives, and judge the degree of collinearity. In order to cope with the huge and complex genomes of Triticeae, strategies have been developed to reduce genome complexity prior to sequencing and mapping, such as exome capture and chromosome genomics. The advances in DNA sequencing technologies make it possible to develop powerful and high-throughput DNA marker technologies such as SNP, DArT and KASPAR, which are suitable for development of markers linked tightly to traits of interest, large-scale screening of progenies of wild hybrids and support production of lines with the introgressed genes of interest and minimum of unwanted chromatin.

Altogether these advances provide a toolbox to develop wheat lines enriched for gene(s) of interest with the smallest amount of undesired alien chromatin. At the same time, it is obvious that we are still at the beginning of what one day may become a routine transfer of alien genes to wheat by interspecific hybridization. In fact, there is another potential obstacle, which so far has received little attention, and that is the genome biology. Almost nothing is known on the behavior of introgressed chromosomes, chromosome segments and/or minute amounts of alien chromatin introgressed into the wheat genome. It is not clear how the wheat genome interacts with introgressed genes and how it influences their function. At the same time, it is important to understand if and how the alien DNA affects the function of the recipient wheat genome. There is an urgent need to clarify the interaction between the host and alien genomes to avoid failed attempts. Luckily, the recent advances in genomics, transcriptomics, epigenomics, proteomics, as well as in cytogenetics, and the analysis of 3D organization of interphase nuclei in particular, are promising to deliver the much needed insights.

Acknowledgments We thank our colleagues Michael Abrouk, Veronika Burešová, Petr Cápal and Gabriella Linc for useful comments and sharing their unpublished results. This work has been supported by the National Program of Sustainability (award no. LO 2014), the Czech Science Foundation (award no. P501-12-G090), the Hungarian National Research Fund (K112226 and K116277), János Bólyai Research Scholarship from the Hungarian Academy of Sciences, and an OECD fellowship (TAD/CRP JA00079297).

References

Adonina IG, Salina EA, Pestsova EG, Röder MS (2005) Transferability of wheat microsatellites to diploid *Aegilops* species and determination of chromosomal localizations of microsatellites in the S genome. Genome 48:959–970

Alheit KV, Busemeyer L, Liu W, Maurer HP, Gowda M, Hahn V, Weissmann S, Ruckelshausen A, Reif JC, Würschum T (2014) Multiple-line cross QTL mapping for biomass yield and plant height in triticale (× *Triticosecale* Wittmack). Theor Appl Genet 127:251–260

Badaeva ED, Friebe B, Gill BS (1996a) Genome differentiation in *Aegilops* 1 Distribution of highly repetitive DNA sequences on chromosomes of diploid species. Genome 39:293–306

Badaeva ED, Friebe B, Gill BS (1996b) Genome differentiation in *Aegilops* 2 Physical mapping of 5S and 18S–26S ribosomal RNA gene families in diploid species. Genome 39:1150–1158

Bandopadhyay R, Sharma S, Rustgi S, Singh R, Kumar A, Balyan HS, Gupta PK (2004) DNA polymorphism among 18 species of *Triticum–Aegilops* complex using wheat EST–SSRs. Plant Sci 166:349–356

Bartoš J, Paux E, Kofler R, Havránková M, Kopecký D, Suchánková P, Šafář J, Šimková H, Town CD, Lelley T, Feuillet C, Doležel J (2008) A first survey of the rye (*Secale cereale*) genome composition through BAC end sequencing of the short arm of chromosome 1R. BMC Plant Biol 8:95

Bedbrook JR, Jones J, O'Dell M, Thompson RD, Flavell RB (1980) A molecular description of telomeric heterochromatin in *Secale* species. Cell 19:545–560

Bennett MD, Finch RA, Barclay IR (1976) The time rate and mechanism of chromosome elimination in *Hordeum* hybrids. Chromosoma 54:175–200

Bentley AR, Scutari M, Gosman N, Faure S, Bedford F, Howell P, Cockram J, Rose GA, Barber T, Irigoyen J, Horsnell R, Pumfrey C, Winnie E, Schacht J, Beauchêne K, Praud S, Greenland A, Balding D, Mackay IJ (2014) Applying association mapping and genomic selection to the dissection of key traits in elite European wheat. Theor Appl Genet 127:2619–2633

Bento M, Gustafson P, Viegas W, Silva M (2010) Genome merger: from sequence rearrangements in triticale to their elimination in wheat–rye addition lines. Theor Appl Genet 121:489–497

Bento M, Pereira HS, Rocheta M, Gustafson P, Viegas W, Silva M (2008) Polyploidization as a retraction force in plant genome evolution: sequence rearrangements in triticale. PLoS One 3, e1402

Berkman PJ, Skarshewski A, Manoli S, Lorenc MT, Stiller J, Smits L, Lai K, Campbell E, Kubaláková M, Šimková H, Batley J, Doležel J, Hernandez P, Edwards D (2011) Sequencing wheat chromosome arm 7BS delimits the 7BS/4AL translocation and reveals homoeologous gene conservation. Theor Appl Genet 124:423–432

Blakeslee A (1937) Redoublement du nombre de chromosomes chez les plantes par traitement chimique. Compt Rend Acad Sci Paris 205:476–479

Del Blanco IA, Rajaram S, Kronstad WE (2001) Agronomic potential of synthetic hexaploid wheat-derived populations. Crop Sci 41:670

Bolibok-Brągoszewska H, Targońska M, Bolibok L, Kilian A, Rakoczy-Trojanowska M (2014) Genome-wide characterization of genetic diversity and population structure in *Secale*. BMC Plant Biol 14:184

Bougas B, Normandeau E, Audet C, Bernatchez L (2013) Linking transcriptomic and genomic variation to growth in brook charr hybrids (*Salvelinus fontinalis* Mitchill.). Heredity 110:492–500

Brenchley R, Spannagl M, Pfeifer M, Barker GLA, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo M-C, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KFX, Edwards KJ, Bevan MW, Hall N (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491:705–710

Busch W, Martin R, Herrmann RG, Hohmann U (1995) Repeated DNA sequences isolated by microdissection. I. Karyotyping of barley (*Hordeum vulgare* L). Genome 38:1082–1090

De Bustos A, Cuadrado A, Soler C, Jouve N (1996) Physical mapping of repetitive DNA sequences and 5S and 18S-26S rDNA in five wild species of the genus *Hordeum*. Chromosome Res 4:491–499

Cabral AL, Jordan MC, McCartney CA, You FM, Humphreys DG, MacLachlan R, Pozniak CJ (2014) Identification of candidate genes, regions and markers for pre-harvest sprouting resistance in wheat (*Triticum aestivum* L). BMC Plant Biol 14:340

Castillo A, Atienza SG, Martín AC (2014) Fertility of CMS wheat is restored by two *Rf* loci located on a recombined acrocentric chromosome. J Exp Bot 65:6667–6677

Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, Forrest K, Saintenac C, Brown-Guedira GL, Akhunova A, See D, Bai G, Pumphrey M, Tomar L, Wong D, Kong S, Reynolds M, da Silva ML, Bockelman H, Talbert L, Anderson JA, Dreisigacker S, Baenziger S, Carter A, Korzun V, Morrell PL, Dubcovsky J, Morell MK, Sorrells ME, Hayden MJ, Akhunov E (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc Natl Acad Sci 110:8057–8062

Cenci A, D'Ovidio R, Tanzarella OA, Ceoloni C, Porceddu E (1999) Identification of molecular markers linked to *Pm13*, an *Aegilops longissima* gene conferring resistance to powdery mildew in wheat. Theor Appl Genet 98:448–454

Chen G, Zheng Q, Bao Y, Liu S, Wang H, Li X (2012) Molecular cytogenetic identification of a novel dwarf wheat line with introgressed *Thinopyrum ponticum* chromatin. J Biosci 37:149–155

Chen P, Liu W, Yuan J, Wang X, Zhou B, Wang S, Zhang S, Feng Y, Yang B, Liu G, Liu D, Qi L, Zhang P, Friebe B, Gill BS (2005) Development and characterization of wheat-*Leymus racemosus* translocation lines with resistance to Fusarium Head Blight. Theor Appl Genet 111:941–948

Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, Pingault L, Sourdille P, Couloux A, Paux E, Leroy P, Mangenot S, Guilhot N, Gouis JL, Balfourier F, Alaux M, Jamilloux V, Poulain J, Durand C, Bellec A, Gaspin C, Safar J, Dolezel J, Rogers J, Vandepoele K, Aury J-M, Mayer K, Berges H, Quesneville H, Wincker P, Feuillet C (2014) Structural and functional partitioning of bread wheat chromosome 3B. Science 345:1249721

Colasuonno P, Maria MA, Blanco A, Gadaleta A (2013) Description of durum wheat linkage map and comparative sequence analysis of wheat mapped DArT markers with rice and *Brachypodium* genomes. BMC Genet 14:114

Comai L (2000) Genetic and epigenetic interactions in allopolyploid plants. Plant Mol Biol 43:387–399

Comai L, Madlung A, Josefsson C, Tyagi A (2003) Do the different parental "heteromes" cause genomic shock in newly formed allopolyploids? Philos Trans R Soc Lond B Biol Sci 358:1149–1155

Cseh A, Kruppa K, Molnár I, Rakszegi M, Doležel J, Molnár-Láng M (2011) Characterization of a new 4BS.7HL wheat/barley translocation line using GISH, FISH and SSR markers and its effect on the β-glucan content of wheat. Genome 54:795–804

Cuadrado A, Cardoso M, Jouve N (2008) Physical organisation of simple sequence repeats (SSRs) in Triticeae: structural, functional and evolutionary implications. Cytogenet Genome Res 120:210–219

Danilova TV, Friebe B, Gill BS (2014) Development of a wheat single gene FISH map for analyzing homoeologous relationship and chromosomal rearrangements within the Triticeae. Theor Appl Genet 127:715–730

Darkó É, Janda T, Majláth I, Szopkó D, Dulai S, Molnár I, Molnár-Láng M (2015) Salt stress response of wheat/barley addition lines carrying chromosomes from the winter barley "Manas". Euphytica 203:491–504

Devos KM, Atkinson MD, Chinoy CN, Francis HA, Harcourt RL, Koebner RMD, Liu CJ, Masojć P, Xie DX, Gale MD (1993) Chromosomal rearrangements in the rye genome relative to that of wheat. Theor Appl Genet 85:673–680

Devos KM, Gale MD (1993) Extended genetic maps of the homoeologous group 3 chromosomes of wheat, rye and barley. Theor Appl Genet 85:649–652

Doležel J, Číhalíková J, Lucretti S (1992) A high-yield procedure for isolation of metaphase chromosomes from root tips of *Vicia faba* L. Planta 188:93–98

Doležel J, Kubaláková M, Paux E, Bartoš J, Feuillet C (2007) Chromosome-based genomics in the cereals. Chromosome Res 15:51–66

Doležel J, Vrána J, Cápal P, Kubaláková M, Burešová V, Šimková H (2014) Advances in plant chromosome genomics. Biotechnol Adv 32:122–136

Dulai S, Molnár I, Haló B, Molnár-Láng M (2010) Photosynthesis in the 7H Asakaze komugi/Manas wheat/barley addition line during salt stress. Acta Agron Hung 58:367–376

Dulai S, Molnár I, Szopkó D, Darkó É, Vojtkó A, Sass-Gyarmati A, Molnár-Láng M (2014) Wheat-*Aegilops biuncialis* amphiploids have efficient photosynthesis and biomass production during osmotic stress. J Plant Physiol 171:509–517

Ellis THN, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ (1998) Polymorphism of insertion sites of *Ty1*-copia class retrotransposons and its use for linkage and diversity analysis in pea. Mol Gen Genet 260:9–19

Endo TR, Kubaláková M, Vrána J, Doležel J (2014) Hyperexpansion of wheat chromosomes sorted by flow cytometry. Genes Genet Syst 89:181–185

Ersfeld K (2004) Fiber-FISH: fluorescence *in situ* hybridization on stretched DNA. Methods Mol Biol 270:395–402

Farkas A, Molnár I, Dulai S, Rapi S, Oldal V, Cseh A, Kruppa K, Molnár-Láng M (2014) Increased micronutrient content (Zn, Mn) in the $3M^b$(4B) wheat-*Aegilops biuncialis* substitution and $3M^b$.4BS translocation identified by GISH and FISH. Genome 57:61–67

Farrer W (1904) Some notes on the wheat "Bobs"; its peculiarities, economic value and origin. Agric Gaz NSW 15:849–854

Fatih AM (1983) Analysis of the breeding potential of wheat-*Agropyron* and wheat-*Elymus* derivatives. I. Agronomic and quality characteristics. Hereditas 98:287–295

Fedak G (1999) Molecular aids for integration of alien chromatin through wide crosses. Genome 42:584–591

Feldman M, Liu B, Segal G, Abbo S, Levy AA, Vega JM (1997) Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. Genetics 147:1381–1387

Feuillet C, Langridge P, Waugh R (2008) Cereal breeding takes a walk on the wild side. Trends Genet 24:24–32

Francki MG, Crasta OR, Sharma HC, Ohm HW, Anderson JM (1997) Structural organization of an alien *Thinopyrum intermedium* group 7 chromosome in US soft red winter wheat (*Triticum aestivum* L). Genome 40:716–722

Friebe B, Jiang J, Raupp WJ, McIntosh RA, Gill BS (1996) Characterization of wheat-alien translocations conferring resistance to diseases and pests: current status. Euphytica 91:59–87

Friebe B, Larter EN (1988) Identification of a complete set of isogenic wheat/rye D-genome substitution lines by means of Giemsa C-banding. Theor Appl Genet 76:473–479

Friebe B, Zeller FJ, Mukai Y, Forster BP, Bartos P, McIntosh RA (1992) Characterization of rust-resistant wheat-*Agropyron intermedium* derivatives by C-banding, *in situ* hybridization and isozyme analysis. Theor Appl Genet 83:775–782

Fu S, Yang M, Fei Y, Tan F, Ren Z, Yan B, Zhang H, Tang Z (2013) Alterations and abnormal mitosis of wheat chromosomes induced by wheat-rye monosomic addition lines. PLoS One 8, e70483

Gerlach WL, Bedbrook JR (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res 7:1869–1885

Gibcus JH, Dekker J (2013) The hierarchy of the 3D genome. Mol Cell 49:773–782

Gill BS, Friebe BR, White FF (2011) Alien introgressions represent a rich source of genes for crop improvement. Proc Natl Acad Sci 108:7657–7658

Gill BS, Kimber G (1974) The giemsa C-banded karyotype of rye. Proc Natl Acad Sci U S A 71:1247–1249

Gill KS, Arumuganathan K, Lee J-H (1999) Isolating individual wheat (*Triticum aestivum*) chromosome arms by flow cytometric analysis of ditelosomic lines. Theor Appl Genet 98:1248–1252

Van Ginkel M, Ogbonnaya F (2008) Using synthetic wheats to breed cultivars better adapted to changing production conditions. Field Crop Res 104:86–94

Giorgi D, Farina A, Grosso V, Gennaro A, Ceoloni C, Lucretti S (2013) FISHIS: fluorescence *in situ* hybridization in suspension and chromosome flow sorting made easy. PLoS One 8, e57994

Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G (2006) Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. Nature 439:749–752

Grob S, Schmid MW, Grossniklaus U (2014) Hi-C analysis in *Arabidopsis* identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila. Mol Cell 55:678–693

Grosso V, Farina A, Gennaro A, Giorgi D, Lucretti S (2012) Flow sorting and molecular cytogenetic identification of individual chromosomes of *Dasypyrum villosum* L. (*H. villosa*) by a single DNA probe. PLoS One 7, e50151

Guadagnuolo R, Bianchi DS, Felber F (2001) Specific genetic markers for wheat, spelt, and four wild relatives: comparison of isozymes, RAPDs, and wheat microsatellites. Genome 44:610–621

Gupta PK, Mir RR, Mohan A, Kumar J (2008) Wheat genomics: present status and future prospects. Int J Plant Genomics 2008:896451

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genomics 270:315–323

Gupta PK, Varshney RK, Sharma PC, Ramesh B (1999) Molecular markers and their applications in wheat breeding. Plant Breeding 118:369–390

Hajjar R, Hodgkin T (2007) The use of wild relatives in crop improvement: a survey of developments over the last 20 years. Euphytica 156:1–13

Han FP, Fedak G, Ouellet T, Liu B (2003) Rapid genomic changes in interspecific and intergeneric hybrids and allopolyploids of Triticeae. Genome 46:716–723

Hart GE, Islam AKMR, Shepherd KW (1980) Use of isozymes as chromosome markers in the isolation and characterization of wheat-barley chromosome addition lines. Genet Res 36:311–325

Haseneyer G, Schmutzer T, Seidel M, Zhou R, Mascher M, Schön C-C, Taudien S, Scholz U, Stein N, Mayer KFX, Bauer E (2011) From RNA-seq to large-scale genotyping - genomics resources for rye (*Secale cereale* L). BMC Plant Biol 11:131

He C, Holme J, Anthony J (2014) SNP genotyping: the KASP assay. Crop Breed (Springer, New York) 2014:75–86

Helguera M, Rivarola M, Clavijo B, Martis MM, Vanzetti LS, González S, Garbus I, Leroy P, Šimková H, Valárik M, Caccamo M, Doležel J, Mayer KFX, Feuillet C, Tranquilli G, Paniego N, Echenique V (2015) New insights into the wheat chromosome 4D structure and virtual gene order, revealed by survey pyrosequencing. Plant Sci 233:200–212

Hernandez P, Martis M, Dorado G, Gálvez S, Pfeifer M, Schaaf S, Jouve N, Šimková H, Valárik M, Doležel J, Mayer KFX (2012) Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. Plant J 69:377–386

Hernández P, Rubio MJ, Martin A (1996) Development of RAPD markers in tritordeum and addition lines of *Hordeum chilense* in *Triticum aestivum*. Plant Breeding 115:52–56

Hurni S, Brunner S, Stirnweis D, Herren G, Peditto D, McIntosh RA, Keller B (2014) The powdery mildew resistance gene *Pm8* derived from rye is suppressed by its wheat ortholog *Pm3*. Plant J 79:904–913

Iehisa JCM, Ohno R, Kimura T, Enoki H, Nishimura S, Okamoto Y, Nasuda S, Takumi S (2014) A high-density genetic map with array-based markers facilitates structural and quantitative trait locus analyses of the common wheat genome. DNA Res 21:555–567

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Islam AKMR, Shepherd KW, Sparrow DHB (1981) Isolation and characterization of euplasmic wheat-barley chromosome addition lines. Heredity 46:161–174

The International Wheat Genome Sequencing Consortium (IWGSC) (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity Arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29:e25

Jackson SA, Wang ML, Goodman HM, Jiang J (1998) Application of fiber-FISH in physical mapping of *Arabidopsis thaliana*. Genome 41:566–572

Jakobson I, Peusha H, Timofejeva L, Järve K (2006) Adult plant and seedling resistance to powdery mildew in a *Triticum aestivum* x *Triticum militinae* hybrid line. Theor Appl Genet 112:760–769

Jakobson I, Reis D, Tiidema A, Peusha H, Timofejeva L, Valárik M, Kladivová M, Šimková H, Doležel J, Järve K (2012) Fine mapping, phenotypic characterization and validation of non-race-specific resistance to powdery mildew in a wheat–*Triticum militinae* introgression line. Theor Appl Genet 125:609–623

Jeffrey Chen Z, Ni Z (2006) Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. Bioessays 28:240–252

Jiang J, Friebe B, Gill B (1993) Recent advances in alien gene transfer in wheat. Euphytica 73:199–212

Jighly A, Oyiga BC, Makdis F, Nazari K, Youssef O, Tadesse W, Abdalla O, Ogbonnaya FC (2015) Genome-wide DArT and SNP scan for QTL associated with resistance to stripe rust (*Puccinia striiformis* f. sp. *tritici*) in elite ICARDA wheat (*Triticum aestivum* L.) germplasm. Theor Appl Genet 128:1277–1295

Jing H-C, Bayon C, Kanyuka K, Berry S, Wenzl P, Huttner E, Kilian A, Hammond-Kosack KE (2009) DArT markers: diversity analyses, genomes comparison, mapping and integration with SSR markers in *Triticum monococcum*. BMC Genomics 10:458

Jing HC, Kornyukhin D, Kanyuka K, Orford S, Zlatska A, Mitrofanova OP, Koebner R, Hammond-Kosack K (2007) Identification of variation in adaptively important traits and genome-wide analysis of trait-marker associations in *Triticum monococcum*. J Exp Bot 58:3749–3764

Jupe J, Stam R, Howden AJ, Morris JA, Zhang R, Hedley PE, Huitema E (2013) *Phytophthora capsici*-tomato interaction features dramatic shifts in gene expression associated with a hemibiotrophic lifestyle. Genome Biol 14:R63

Kalih R, Maurer HP, Miedaner T (2015) Genetic architecture of fusarium head blight resistance in four winter triticale populations. Phytopathology 105:334–341

Karafiátová M, Bartoš J, Kopecký D, Ma L, Sato K, Houben A, Stein N, Doležel J (2013) Mapping non-recombining regions in barley using multicolor FISH. Chromosome Res 21:739–751

Kashkush K, Feldman M, Levy AA (2002) Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics 160:1651–1659

Kerber ER, Aung T (1999) Leaf rust resistance gene *Lr34* associated with nonsuppression of stem rust resistance in the wheat cultivar Canthatch. Phytopathology 89:518–521

Kerber ER, Dyck PL (1973) Inheritance of stem rust resistance transferred from diploid wheat (*Triticum monococcum*) to tetraploid and hexaploid wheat and chromosome location of the gene involved. Can J Genet Cytol 15:397–409

Khalil HB, Ehdaeivand M-R, Xu Y, Laroche A, Gulick PJ (2015) Identification and characterization of rye genes not expressed in allohexaploid triticale. BMC Genomics 16:281

Kihara H (1937) Genomanalyse bei *Triticum* und *Aegilops*. Kurze übersicht über die Ergebnisse der Jahre 1934–36. Mem Coll Agr Kyoto Imp Univ 41:1–61

King J, Armstead I, Harper J, Ramsey L, Snape J, Waugh R, James C, Thomas A, Gasior D, Kelly R, Roberts L, Gustafson P, King I (2013) Exploitation of interspecific diversity for monocot crop improvement. Heredity 110:475–483

King IP, Purdie KA, Rezanoor HN, Koebner RMD, Miller TE, Reader SM, Nicholson P (1993) Characterization of *Thinopyrum bessarabicum* chromosome segments in wheat using random amplified polymorphic DNAs (RAPDs) and genomic in situ hybridization. Theor Appl Genet 86:895–900

Koba T, Takumi S, Shimada T (1997) Isolation, identification and characterization of disomic and translocated barley chromosome addition lines of common wheat. Euphytica 96:289–296

Kroupin PY, Divashuk MG, Fesenko IA, Karlov GI (2012) Evaluating wheat microsatellite markers for the use in genetic analysis of *Thinopyrum*, *Dasypyrum*, and *Pseudoroegneria* species. Dataset Paper Sci 2013, e949637

Kruppa K, Sepsi A, Szakács É, Röder MS, Molnár-Láng M (2013) Characterization of a 5HS-7DS.7DL wheat-barley translocation line and physical mapping of the 7D chromosome using SSR markers. J Appl Genet 54:251–258

Kruse A (1973) *Hordeum*×*Triticum* hybrids. Hereditas 73:157–161

Kubaláková M, Kovářová P, Suchánková P, Číhalíková J, Bartoš J, Lucretti S, Watanabe N, Kianian SF, Doležel J (2005) Chromosome sorting in tetraploid wheat and its potential for genome analysis. Genetics 170:823–829

Kubaláková M, Valárik M, Barto J, Vrána J, Číhalíková J, Molnár-Láng M, Dolezel J (2003) Analysis and sorting of rye (*Secale cereale* L) chromosomes using flow cytometry. Genome 46:893–905

Kubaláková M, Vrána J, Číhalíková J, Šimková H, Doležel J (2002) Flow karyotyping and chromosome sorting in bread wheat (*Triticum aestivum* L). Theor Appl Genet 104:1362–1372

Kuraparthy V, Chhuneja P, Dhaliwal HS, Kaur S, Bowden RL, Gill BS (2007) Characterization and mapping of cryptic alien introgression from *Aegilops geniculata* with new leaf rust and stripe rust resistance genes *Lr57* and *Yr40* in wheat. Theor Appl Genet 114:1379–1389

Langridge P, Lagudah ES, Holton TA, Appels R, Sharp PJ, Chalmers KJ (2001) Trends in genetic and genome analyses in wheat: a review. Aust J Agric Res 52:1043–1077

Lapitan NLV, Sears RG, Gill BS (1984) Translocations and other karyotypic structural changes in wheat x rye hybrids regenerated from tissue culture. Theor Appl Genet 68:547–554

Laurie DA, Bennett MD (1986) Wheat×maize hybridization. Can J Genet Cytol 28:313–316

Lee H-S, Chen ZJ (2001) Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. Proc Natl Acad Sci U S A 98:6753–6758

Lee J-H, Arumuganathan K, Chung Y-S, Kim K-Y, Chung W-B, Bae K-S, Kim D-H, Chung D-S, Kwon O-C (2000) Flow cytometric analysis and chromosome sorting of barley (*Hordeum vulgare* L). Mol Cells 10:619–625

Lee J-H, Arumuganathan K, Yen Y, Kaeppler S, Kaeppler H, Baenziger PS (1997) Root tip cell cycle synchronization and metaphase-chromosome isolation suitable for flow sorting in common wheat (*Triticum aestivum* L). Genome 40:633–638

Levy AA, Feldman M (2004) Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. Biol J Linn Soc 82:607–613

Li J, Wan H-S, Yang W-Y (2014) Synthetic hexaploid wheat enhances variation and adaptive evolution of bread wheat in breeding processes. J Syst Evol 52:735–742

Li LJ, Arumuganathan K, Rines HW, Phillips RL, Riera-Lizarazu O, Sandhu D, Zhou Y, Gill KS (2001) Flow cytometric sorting of maize chromosome 9 from an oat-maize chromosome addition line. Theor Appl Genet 102:658–663

Linc G, Sepsi A, Molnár-Láng M (2012) A FISH karyotype to study chromosome polymorphisms for the *Elytrigia elongata* E genome. Cytogenet Genome Res 136:138–144

Ling H-Q, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y, Gao C, Wu H, Li Y, Cui Y, Guo X, Zheng S, Wang B, Yu K, Liang Q, Yang W, Lou X, Chen J, Feng M, Jian J, Zhang X, Luo G, Jiang Y, Liu J, Wang Z, Sha Y, Zhang B, Wu H, Tang D, Shen Q, Xue P, Zou S, Wang X, Liu X, Wang F, Yang Y, An X, Dong Z, Zhang K, Zhang X, Luo M-C, Dvorak J, Tong Y, Wang J, Yang H, Li Z, Wang D, Zhang A, Wang J (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496:87–90

Liu B, Vega JM, Segal G, Abbo S, Rodova M, Feldman M (1998) Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. I. Changes in low-copy noncoding DNA sequences. Genome 41:272–277

Liu W-H, Luan Y, Wang J-C, Wang X-G, Su J-J, Zhang J-P, Yang X-M, Gao A-N, Li L-H (2010) Production and identification of wheat – *Agropyron cristatum* (1.4P) alien translocation lines. Genome 53:472–481

Lucas SJ, Akpınar BA, Šimková H, Kubaláková M, Doležel J, Budak H (2014) Next-generation sequencing of flow-sorted wheat chromosome 5D reveals lineage-specific translocations and widespread gene duplications. BMC Genomics 15:1080

Lukaszewski AJ, Gustafson JP (1983) Translocations and modifications of chromosomes in triticale×wheat hybrids. Theor Appl Genet 64:239–248

Luo M-C, Gu YQ, You FM, Deal KR, Ma Y, Hu Y, Huo N, Wang Y, Wang J, Chen S, Jorgensen CM, Zhang Y, McGuire PE, Pasternak S, Stein JC, Ware D, Kramer M, McCombie WR,

Kianian SF, Martis MM, Mayer KFX, Sehgal SK, Li W, Gill BS, Bevan MW, Šimková H, Doležel J, Weining S, Lazo GR, Anderson OD, Dvorak J (2013) A 4-gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor. Proc Natl Acad Sci U S A 110:7940–7945

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

Lysák MA, Číhalíková J, Kubaláková M, Simková H, Künzel G, Dolezel J (1999) Flow karyotyping and sorting of mitotic chromosomes of barley (*Hordeum vulgare* L). Chromosome Res 7:431–444

Martis MM, Zhou R, Haseneyer G, Schmutzer T, Vrána J, Kubaláková M, König S, Kugler KG, Scholz U, Hackauf B, Korzun V, Schön C-C, Doležel J, Bauer E, Mayer KFX, Stein N (2013) Reticulate evolution of the rye genome. Plant Cell 25:3685–3698

Ma X-F, Gustafson JP (2008) Allopolyploidization-accommodated genomic sequence changes in triticale. Ann Bot 101:825–832

Ma X-F, Gustafson JP (2006) Timing and rate of genome variation in triticale following allopolyploidization. Genome 49:950–958

Mayer KFX, Martis M, Hedley PE, Šimková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubaláková M, Suchánková P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Dolezel J, Waugh R, Stein N (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263

Mayer KFX, Taudien S, Martis M, Šimková H, Suchánková P, Gundlach H, Wicker T, Petzold A, Felder M, Steuernagel B, Scholz U, Graner A, Platzer M, Doležel J, Stein N (2009) Gene content and virtual gene order of barley chromosome 1H. Plant Physiol 151:496–505

McArthur RI, Zhu X, Oliver RE, Klindworth DL, Xu SS, Stack RW, Wang RR-C, Cai X (2012) Homoeology of *Thinopyrum junceum* and *Elymus rectisetus* chromosomes to wheat and disease resistance conferred by the *Thinopyrum* and *Elymus* chromosomes in wheat. Chromosome Res 20:699–715

McIntyre CL, Pereira S, Moran LB, Appels R (1990) New *Secale cereale* (rye) DNA derivatives for the detection of rye chromosome segments in wheat. Genome 33:635–640

Meurant G (1982) Advances in genetics. Academic, New York, NY

Mohan A, Goyal A, Singh R, Balyan HS, Gupta PK (2007) Physical mapping of wheat and rye expressed sequence tag–simple sequence repeats on wheat chromosomes. Crop Sci 47:3

Molnár I, Gáspár L, Sárvári É, Dulai S, Hoffmann B, Molnár-Láng M, Galiba G (2004) Physiological and morphological responses to water stress in *Aegilops biuncialis* and *Triticum aestivum* genotypes with differing tolerance to drought. Funct Plant Biol 31:1149–1159

Molnár I, Benavente E, Molnár-Láng M (2009) Detection of intergenomic chromosome rearrangements in irradiated *Triticum aestivum-Aegilops biuncialis* amphiploids by multicolour genomic *in situ* hybridization. Genome 52:156–165

Molnár I, Cifuentes M, Schneider A, Benavente E, Molnár-Láng M (2011a) Association between simple sequence repeat-rich chromosome regions and intergenomic translocation breakpoints in natural populations of allopolyploid wild wheats. Ann Bot 107:65–76

Molnár I, Kubaláková M, Šimková H, Cseh A, Molnár-Láng M, Doležel J (2011b) Chromosome isolation by flow sorting in *Aegilops umbellulata* and *Ae. comosa* and their allotetraploid hybrids *Ae. biuncialis* and *Ae. geniculata*. PLoS One 6, e27708

Molnár I, Kubaláková M, Šimková H, Farkas A, Cseh A, Megyeri M, Vrána J, Molnár-Láng M, Doležel J (2014) Flow cytometric chromosome sorting from diploid progenitors of bread wheat, *T. urartu*, *Ae. speltoides* and *Ae. tauschii*. Theor Appl Genet 127:1091–1104

Molnár I, Šimková H, Leverington-Waite M, Goram R, Cseh A, Vrána J, Farkas A, Doležel J, Molnár-Láng M, Griffiths S (2013) Syntenic relationships between the U and M genomes of *Aegilops*, wheat and the model species *Brachypodium* and rice as revealed by COS markers. PLoS One 8, e70844

Molnár I, Vrána J, Farkas A, Kubaláková M, Cseh A, Molnár-Láng M, Doležel J (2015) Flow sorting of C-genome chromosomes from wild relatives of wheat *Aegilops markgrafii*, *Ae triuncia-*

lis and *Ae cylindrica*, and their molecular organization. Ann Bot 116:189–200. doi:10.1093/aob/mcv073

Molnár-Láng M, Kruppa K, Cseh A, Bucsi J, Linc G (2012) Identification and phenotypic description of new wheat: six-rowed winter barley disomic additions. Genome 55:302–311

Molnár-Láng M, Linc G, Friebe BR, Sutka J (2000) Detection of wheat-barley translocations by genomic *in situ* hybridization in derivatives of hybrids multiplied *in vitro*. Euphytica 112:117–123

Molnár-Láng M, Novotny C, Linc G, Nagy ED (2005) Changes in the meiotic pairing behaviour of a winter wheat-winter barley hybrid maintained for a long term in tissue culture, and tracing the barley chromatin in the progeny using GISH and SSR markers. Plant Breeding 124:247–252

Molnár-Láng M, Molnár I, Szakács É, Linc G, Bedö Z (2014) Production and molecular cytogenetic identification of wheat-alien hybrids and introgression lines. In: Tuberosa R, Graner A, Frison E (eds) Genomics of plant genetic resources. Springer, Dordrecht, pp 255–283, doi:10.1007/978-94-007-7572-5

Montilla-Bascón G, Rispail N, Sánchez-Martín J, Rubiales D, Mur LAJ, Langdon T, Howarth CJ, Prats E (2015) Genome-wide association study for crown rust (*Puccinia coronata* f.sp. *avenae*) and powdery mildew (*Blumeria graminis* f.sp. *avenae*) resistance in an oat (*Avena sativa*) collection of commercial varieties and landraces. Front. Plant Sci 6:103

Mujeeb-Kazi A (1995) Intergeneric crosses: hybrid production and utilization. In: Mujeeb-Kazi A, Hettel GP (eds) Utilizing wild grass biodiversity in wheat improvement: 15 years of wide cross research at CIMMYT. CIMMYT, Mexico, p 140

Mukhopadhyay P, Singla-Pareek SL, Reddy MK, Sopory SK (2013) Stress-mediated alterations in chromatin architecture correlate with down-regulation of a gene encoding 60S rpL32 in rice. Plant Cell Physiol 54(4):528–540

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plant 15:473–497

Nagaki K, Tsujimoto H, Isono K, Sasakuma T (1995) Molecular characterization of a tandem repeat, *Afa* family, and its distribution among Triticeae. Genome 38:479–486

Nagy ED, Lelley T (2003) Genetic and physical mapping of sequence-specific amplified polymorphic (SSAP) markers on the 1RS chromosome arm of rye in a wheat background. Theor Appl Genet 107:1271–1277

Nagy ED, Molnár I, Schneider A, Kovács G, Molnár-Láng M (2006) Characterization of chromosome-specific S-SAP markers and their use in studying genetic diversity in *Aegilops* species. Genome 49:289–296

Nagy ED, Molnár-Láng M, Linc G, Láng L (2002) Identification of wheat-barley translocations by sequential GISH and two-colour FISH in combination with the use of genetically mapped barley SSR markers. Genome 45:1238–1247

Ng DW-K, Zhang C, Miller M, Shen Z, Briggs SP, Chen ZJ (2012) Proteomic divergence in *Arabidopsis* autopolyploids and allopolyploids and their progenitors. Heredity 108:419–430

Osborn TC, Chris Pires J, Birchler JA, Auger DL, Jeffery Chen Z, Lee H-S, Comai L, Madlung A, Doerge RW, Colot V, Martienssen RA (2003) Understanding mechanisms of novel gene expression in polyploids. Trends Genet 19:141–147

Pace CD, Snidaro D, Ciaffi M, Vittori D, Ciofo A, Cenci A, Tanzarella OA, Qualset CO, Mugnozza GTS (2001) Introgression of *Dasypyrum villosum* chromatin into common wheat improves grain protein quality. Euphytica 117:67–75

Parisod C, Senerchia N (2012) Responses of transposable elements to polyploidy In plant transposable elements. In: Grandbastien M-A, Casacuberta JM (eds) Topics in current genetics. Springer, Berlin, pp 147–168

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-

ur-Rahman, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Peil A, Korzun V, Schubert V, Schumann E, Weber WE, Röder MS (1998) The application of wheat microsatellites to identify disomic *Triticum aestivum-Aegilops markgrafii* addition lines. Theor Appl Genet 96:138–146

Peleg Z, Saranga Y, Suprunova T, Ronin Y, Röder MS, Kilian A, Korol AB, Fahima T (2008) High-density genetic map of durum wheat x wild emmer wheat based on SSR and DArT markers. Theor Appl Genet 117:103–115

Petersen S, Lyerly JH, Worthington ML, Parks WR, Cowger C, Marshall DS, Brown-Guedira G, Murphy JP (2015) Mapping of powdery mildew resistance gene *Pm53* introgressed from *Aegilops speltoides* into soft red winter wheat. Theor Appl Genet 128:303–312

Pradhan GP, Prasad PVV (2015) Evaluation of wheat chromosome translocation lines for high temperature stress tolerance at grain filling stage. PLoS One 10

Pumphrey M, Bai J, Laudencia-Chingcuanco D, Anderson O, Gill BS (2009) Nonadditive expression of homoeologous genes is established upon polyploidization in hexaploid wheat. Genetics 181:1147–1157

Qi L, Cao M, Chen P, Li W, Liu D (1996) Identification, mapping, and application of polymorphic DNA associated with resistance gene *Pm21* of wheat. Genome 39:191–197

Qi L, Echalier B, Friebe B, Gill BS (2003) Molecular characterization of a set of wheat deletion stocks for use in chromosome bin mapping of ESTs. Funct Integr Genomics 3:39–55

Qi L, Friebe B, Zhang P, Gill BS (2007) Homoeologous recombination, chromosome engineering and crop improvement. Chromosome Res 15:3–19

Qi LL, Echalier B, Chao S, Lazo GR, Butler GE, Anderson OD, Akhunov ED, Dvorák J, Linkiewicz AM, Ratnasiri A, Dubcovsky J, Bermudez-Kandianis CE, Greene RA, Kantety R, La Rota CM, Munkvold JD, Sorrells SF, Sorrells ME, Dilbirligi M, Sidhu D, Erayman M, Randhawa HS, Sandhu D, Bondareva SN, Gill KS, Mahmoud AA, Ma XF, Miftahudin, Gustafson JP, Conley EJ, Nduati V, Gonzalez-Hernandez JL, Anderson JA, Peng JH, Lapitan NL, Hossain KG, Kalavacharla V, Kianian SF, Pathan MS, Zhang DS, Nguyen HT, Choi DW, Fenton RD, Close TJ, McGuire PE, Qualset CO, Gill BS (2004) A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. Genetics 168:701–712

Qi LL, Wang SL, Chen PD, Liu DJ, Friebe B, Gill BS (1997) Molecular cytogenetic analysis of *Leymus racemosus* chromosomes added to wheat. Theor Appl Genet 95:1084–1091

Queen RA, Gribbon BM, James C, Jack P, Flavell AJ (2003) Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat. Mol Genet Genomics 271:91–97

Rabinovich SV (1998) Importance of wheat-rye translocations for breeding modern cultivar of *Triticum aestivum* L. Euphytica 100:323–340

Rayburn AL, Gill BS (1985) Use of biotin-labeled probes to map specific DNA sequences on wheat chromosomes. J Hered 76:78–81

Reynolds MP, Calderini DF, Condon AG, Rajaram S (2001) Physiological basis of yield gains in wheat associated with the *Lr19* translocation from *Agropyron elongatum* Wheat in a global environment. In: Bedö Z, Láng L (eds) Developments in plant breeding. Springer, Amsterdam, pp 345–351

Riley R, Chapman V (1958) Genetic control of the cytologically diploid behaviour of hexaploid wheat. Nature 182:713–715

Riley R, Macer RCF (1966) The chromosomal distribution of the genetic resistance of rye to wheat pathogens. Can J Genet Cytol 8:640–653

Ruperao P, Chan C-KK, Azam S, Karafiátová M, Hayashi S, Čížková J, Saxena RK, Šimková H, Song C, Vrána J, Chitikineni A, Visendi P, Gaur PM, Millán T, Singh KB, Taran B, Wang J, Batley J, Doležel J, Varshney RK, Edwards D (2014) A chromosomal genomics approach to assess and validate the desi and kabuli draft chickpea genome assemblies. Plant Biotechnol J 12:778–786

Sanei M, Pickering R, Kumke K, Nasuda S, Houben A (2011) Loss of centromeric histone H3 (*CENH3*) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. Proc Natl Acad Sci U S A 108:E498–E505

Schlegel R, Cakmak I, Torun B, Eker S, Tolay I, Ekiz H, Kalayci M, Braun HJ (1998) Screening for zinc efficiency among wheat relatives and their utilisation for alien gene transfer. Euphytica 100:281–286

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Schneider A, Molnár I, Molnár-Láng M (2008) Utilisation of *Aegilops* (goatgrass) species to widen the genetic diversity of cultivated wheat. Euphytica 163:1–19

Schubert I, Shi F, Fuchs J, Endo TR (1998) An efficient screening for terminal deletions and translocations of barley chromosomes added to common wheat. Plant J 14:489–495

Schubert V, Rudnik R, Schubert I (2014) Chromatin associations in *Arabidopsis* interphase nuclei. Front Genet 5:389

Schwarzacher T, Leitch AR, Bennett MD, Heslop-Harrison JS (1989) *In situ* localization of parental genomes in a wide hybrid. Ann Bot 64:315–324

Schwarzacher T, Wang ML, Leitch AR, Moore G, Heslop-Harrison JS, Miller N (1997) Flow cytometric analysis of the chromosomes and stability of a wheat cell-culture line. Theor Appl Genet 94:91–97

Sears ER (1956) The transfer of leaf-rust resistance from *Aegilops umbellulata* to wheat. Brookhaven symposia in biology, 1956. Genet Plant Breed 1956:1–22

Senerchia N, Felber F, Parisod C (2014) Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. New Phytol 202:975–985

Sepsi A, Molnár I, Szalay D, Molnár-Láng M (2008) Characterization of a leaf rust-resistant wheat-*Thinopyrum ponticum* partial amphiploid BE-1, using sequential multicolor GISH and FISH. Theor Appl Genet 116:825–834

Sergeeva EM, Afonnikov DA, Koltunova MK, Gusev VD, Miroshnichenko LA, Vrána J, Kubaláková M, Poncet C, Sourdille P, Feuillet C, Doležel J, Salina EA (2014) Common wheat chromosome 5B composition analysis using low-coverage 454 sequencing. Plant Genome 7:2

Seyfarth R, Feuillet C, Schachermayr G, Winzeler M, Keller B (1999) Development of a molecular marker for the adult plant leaf rust resistance gene *Lr35* in wheat. Theor Appl Genet 99:554–560

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA (2001) Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell 13:1749–1760

Šimková H, Svensson JT, Condamine P, Hřibová E, Suchánková P, Bhat PR, Bartoš J, Šafář J, Close TJ, Doležel J (2008) Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. BMC Genomics 9:294

Soltis DE, Soltis PS, Tate JA (2004) Advances in the study of polyploidy since plant speciation. New Phytol 161:173–191

Stirnweis D, Milani SD, Brunner S, Herren G, Buchmann G, Peditto D, Jordan T, Keller B (2014) Suppression among alleles encoding nucleotide-binding-leucine-rich repeat resistance proteins interferes with resistance in F1 hybrid and allele-pyramided wheat plants. Plant J 79:893–903

Suchánková P, Kubaláková M, Kovářová P, Bartoš J, Číhalíková J, Molnár-Láng M, Endo TR, Doležel J (2006) Dissection of the nuclear genome of barley by chromosome flow sorting. Theor Appl Genet 113:651–659

Szakács É, Molnár-Láng M (2007) Development and molecular cytogenetic identification of new winter wheat – winter barley ("Martonvásári 9 kr1" – "Igri") disomic addition lines. Genome 50:43–50

Szakács É, Molnár-Láng M (2008) Fluorescent *in situ* hybridization polymorphism on the 1RS chromosome arms of cultivated *Secale cereale* species. Cereal Res Commun 36:247–255

Szakács É, Kruppa K, Molnár-Láng M (2013) Analysis of chromosomal polymorphism in barley (*Hordeum vulgare* L. ssp. *vulgare*) and between *H. vulgare* and *H. chilense* using three-color fluorescence *in situ* hybridization (FISH). J Appl Genet 54:427–433

Tanaka T, Kobayashi F, Joshi GP, Onuki R, Sakai H, Kanamori H, Wu J, Simkova H, Nasuda S, Endo TR, Hayakawa K, Doležel J, Ogihara Y, Itoh T, Matsumoto T, Handa H (2014) Next-generation survey sequencing and the molecular organization of wheat chromosome 6B. DNA Res 21:103–114

Tang KS, Hart GE (1975) Use of isozymes as chromosome markers in wheat-rye addition lines and in triticale. Genet Res 26:187–201

Tautz D (1989) Hypervariabflity of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471

Tayalé A, Parisod C (2013) Natural pathways to polyploidy in plants and consequences for genome reorganization. Cytogenet Genome Res 140:79–96

The International Barley Genome Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716

The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768

Tiwari VK, Wang S, Sehgal S, Vrána J, Friebe B, Kubaláková M, Chhuneja P, Doležel J, Akhunov E, Kalia B, Sabir J, Gill BS (2014) SNP Discovery for mapping alien introgressions in wheat. BMC Genomics 15:273

Valárik M, Bartos J, Kovárová P, Kubaláková M, de Jong JH, Dolezel J (2004) High-resolution FISH on super-stretched flow-sorted plant chromosomes. Plant J 37:940–950

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Vrána J, Kubaláková M, Simková H, Číhalíkováí J, Lysák MA, Dolezel J (2000) Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L). Genetics 156:2033–2041

Wang ML, Leitch AR, Schwarzacher T, Heslop-Harrison JS, Moore G (1992) Construction of a chromosome-enriched *Hpa*II library from flow-sorted wheat chromosomes. Nucleic Acids Res 20:1897–1901

Wang RR-C, Chen J, Joppa LR (1995) Production and identification of chromosome specific RAPD markers for Langdon durum wheat disomic substitution lines. Crop Sci 35:886

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, International Wheat Genome Sequencing Consortium, Lillemo M, Mather D, Appels R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo MC, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol J 12:787–796

Warburton ML, Crossa J, Franco J, Kazi M, Trethowan R, Rajaram S, Pfeiffer W, Zhang P, Dreisigacker S, van Ginkel M (2006) Bringing wild relatives back into the family: recovering genetic diversity in CIMMYT improved wheat germplasm. Euphytica 149:289–301

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT, Powell W (1997) Genetic distribution of *Bare–1*-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Genet 253:687–694

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Wilson AS (1876) Wheat and rye hybrids. Trans Proc Bot Soc (Edinb) 12:286–288

Wulff BBH, Moscou MJ (2014) Strategies for transferring resistance into wheat: from wide crosses to GM cassettes. Front Plant Sci 5:692

Wu Y, Sun Y, Wang X, Lin X, Sun S, Shen K, Wang J, Jiang T, Zhong S, Xu C, Liu B (2015) Transcriptome shock in an interspecific F1 triploid hybrid of *Oryza* revealed by RNA sequencing. J Integr Plant Biol. doi:10.1111/jipb.12357

Xiong LZ, Xu CG, Maroof MAS, Zhang Q (1999) Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique. Mol Gen Genet 261:439–446

Yabe S, Hara T, Ueno M, Enoki H, Kimura T, Nishimura S, Yasui Y, Ohsawa R, Iwata H (2014) Rapid genotyping with DNA micro-arrays for high-density linkage mapping and QTL mapping in common buckwheat (*Fagopyrum esculentum* Moench.). Breed Sci 64:291–299

Yamamoto M, Mukai Y (1989) Application of fluorescence *in situ* hybridization to molecular cytogenetics of wheat. Wheat Inf Serv 69:30–32

Yang X, Cao A, Sun Y, Chen P (2013) Tracing the location of powdery mildew resistance-related gene *Stpk-V* by FISH with a TAC clone in *Triticum aestivum-Haynaldia villosa* alien chromosome lines. Chin Sci Bull 58:4084–4091

Yasui Y, Nasuda S, Matsuoka Y, Kawahara T (2001) The *Au* family, a novel short interspersed element (SINE) from *Aegilops umbellulata*. Theor Appl Genet 102:463–470

Yingshan D, Xiuling B, Yushi L, Mengyuan H, Bao L (2004) Molecular characterization of a cryptic wheat-*Thinopyrum intermedium* translocation line: evidence for genomic instability in nascent allopolyploid and aneuploid lines. Genet Mol Biol 27:237–241

Yoo M-J, Szadkowski E, Wendel JF (2013) Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity 110:171–180

Yu H, Deng Z, Xiang C, Tian J (2014) Analysis of diversity and linkage disequilibrium mapping of agronomic traits on B-genome of wheat. J Genomics 2:20–30

Yu J-K, Rota ML, Kantety RV, Sorrells ME (2004) EST derived SSR markers for comparative mapping in wheat and rice. Mol Genet Genomics 271:742–751

Zarco-Hernandez JA, Santiveri F, Michelena A, Javier Peña R (2005) Durum wheat (*Triticum turgidum* L.) carrying the 1BL.1RS chromosomal translocation: agronomic performance and quality characteristics under Mediterranean conditions. Eur J Agron 22:33–43

Zhang H, Jia J, Gale MD, Devos KM (1998) Relationships between the chromosomes of *Aegilops umbellulata* and wheat. Theor Appl Genet 96:69–75

Zhang J, Liu W, Han H, Song L, Bai L, Gao Z, Zhang Y, Yang X, Li X, Gao A, Li L (2015) De novo transcriptome sequencing of *Agropyron cristatum* to identify available gene resources for the enhancement of wheat. Genomics. doi:10.1016/j.ygeno.2015.04.003

Zhang P, Li W, Fellers J, Friebe B, Gill BS (2004a) BAC-FISH in wheat identifies chromosome landmarks consisting of different types of transposable elements. Chromosoma 112:288–299

Zhang P, Li W, Friebe B, Gill BS (2004b) Simultaneous painting of three genomes in hexaploid wheat by BAC-FISH. Genome 47:979–987

6.2 Appendix II

Transcriptome reprogramming due to the introduction of a barley telosome into bread wheat affects more barley genes than wheat

Elodie Rey, Michael Abrouk, Gabriel Keeble-Gagnère, Miroslava Karafiátová, Jan Vrána, Sandrine Balzergue, Ludivine Soubigou-Taconnat, Véronique Brunaud, Marie-Laure Martin-Magniette, Takashi R. Endo, Jan Bartoš, International Wheat Genome Sequencing Consortium (IWGSC), Rudi Appels and Jaroslav Doležel

Accepted paper

Plant Biotechnology Journal

doi: 10.1111/pbi.12913.

Transcriptome reprogramming due to the introduction of a barley telosome into bread wheat affects more barley genes than wheat

Elodie Rey[1], Michael Abrouk[1], Gabriel Keeble-Gagnère[2], Miroslava Karafiátová[1], Jan Vrána[1], Sandrine Balzergue[3,4,5], Ludivine Soubigou-Taconnat[3,4], Véronique Brunaud[3,4], Marie-Laure Martin-Magniette[3,4,6], Takashi R. Endo[7], Jan Bartoš[1], International Wheat Genome Sequencing Consortium (IWGSC), Rudi Appels[8] and Jaroslav Doležel[1,*]

[1]*Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic*
[2]*AgriBio, Melbourne, Vic., Australia*
[3]*Institute of Plant Sciences Paris Saclay IPS2, CNRS, INRA, Université Paris-Sud, Université Evry, Université Paris-Saclay, Orsay, France*
[4]*Institute of Plant Sciences Paris-Saclay IPS2, Paris Diderot, Sorbonne Paris-Cité, Orsay, France*
[5]*IRHS, Université d'Angers, INRA, AGROCAMPUS-Ouest, SFR4207 QUASAV, Université Bretagne Loire, Beaucouzé, France*
[6]*UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, Paris, France*
[7]*Department of Plant Life Science, Faculty of Agriculture, Ryukoku University, Shiga, Japan*
[8]*Murdoch University, Perth, WA, Australia*

Received 5 December 2017;
revised 17 February 2018;
accepted 24 February 2018.
**Correspondence* (Tel xxxx; fax xxxx; email dolezel@ueb.cas.cz)

Keywords: alien introgression, RNA-seq, gene transcription, transcriptome modification, chromosomal rearrangement, deletion.

Summary

Despite a long history, the production of useful alien introgression lines in wheat remains difficult mainly due to linkage drag and incomplete genetic compensation. In addition, little is known about the molecular mechanisms underlying the impact of foreign chromatin on plant phenotype. Here, a comparison of the transcriptomes of barley, wheat and a wheat–barley 7HL addition line allowed the transcriptional impact both on 7HL genes of a non-native genetic background and on the wheat gene complement as a result of the presence of 7HL to be assessed. Some 42% (389/923) of the 7HL genes assayed were differentially transcribed, which was the case for only 3% (960/35 301) of the wheat gene complement. The absence of any transcript in the addition line of a suite of chromosome 7A genes implied the presence of a 36 Mbp deletion at the distal end of the 7AL arm; this deletion was found to be in common across the full set of Chinese Spring/Betzes barley addition lines. The remaining differentially transcribed wheat genes were distributed across the whole genome. The up-regulated barley genes were mostly located in the proximal part of the 7HL arm, while the down-regulated ones were concentrated in the distal part; as a result, genes encoding basal cellular functions tended to be transcribed, while those encoding specific functions were suppressed. An insight has been gained into gene transcription in an alien introgression line, thereby providing a basis for understanding the interactions between wheat and exotic genes in introgression materials.

Introduction

Bread wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*) are the two leading temperate small-grained cereals, and together supply, directly or indirectly, a major proportion of our calorific requirements. The major breeding priority over many decades has been to increase their productivity, and this goal was growing in importance in the face of a continuing expansion in the size of the human population, changes in the dietary habits and the threat of impending climate change (Curtis and Halford, 2014). A key means to meet this challenge is to extend the genetic diversity available for breeding, and one of the best ways to achieve this is likely to be to unlock the largely untapped diversity harboured by the crops' wild and cultivated relatives (Brozynska *et al.*, 2016; Feuillet *et al.*, 2008; Mondal *et al.*, 2016). In wheat, continuing technological advances have, over many years, improved the success rate of wide hybridization (Blakeslee, 1937; Kruse, 1973; Mujeeb-Kazi, 1995; Murashige and Skoog, 1962), while progress in inducing introgression has succeeded in introducing subchromosomal segments of exotic origin (Endo, 1988, 2007; Griffiths *et al.*, 2006; Jiang *et al.*, 1993; Riley and Chapman, 1958). Many of these can be readily visualized at both the cytogenetic and the genetic levels (Fedak, 1999; Gupta *et al.*, 1999; Rayburn and Gill, 1985; Rey *et al.*, 2015; Yamamoto and Mukai, 1989). A few introgression products have made positive contributions to grain yield, resistance to disease and tolerance of abiotic stress (Börner *et al.*, 2015; Fatih, 1983; Friebe *et al.*, 1996; Mondal *et al.*, 2016; Reynolds *et al.*, 2001; Wulff and Moscou, 2014), but the majority have had little impact on wheat improvement, as their presence

Please cite this article as: Rey, E., Abrouk, M., Keeble-Gagnère, G., Karafiátová, M., Vrána, J., Balzergue, S., Soubigou-Taconnat, L., Brunaud, V., Martin-Magniette, M.-L., Endo, T.R., Bartoš, J., International Wheat Genome Sequencing Consortium (IWGSC), Appels, R. and Doležel, J. (2018) Transcriptome reprogramming due to the introduction of a barley telosome into bread wheat affects more barley genes than wheat. *Plant Biotechnol. J.*, https://doi.org/10.1111/pbi.12913

has been associated with either a yield penalty and/or an end product defect.

While the long-standing assumption has been that the poor performance of introgression lines is due to linkage drag, together with—in the case where an exotic chromatin segment replaces its wheat homeolog—a less than perfect compensation for the loss of wheat genes, there is some evidence that gene interactions between the recipient and the donor genes are also implicated. For example, the suppression of a gene determining resistance to a pathogen has been shown to be caused by its interaction with the recipient genome (Chen *et al.*, 2005; Kerber and Aung, 1999; Kerber and Dyck, 1973), in some cases shown to be the result of post-translational modifications being imposed on the donor gene product (Stirnweis *et al.*, 2014). According to Hurni *et al.* (2014), deleterious effects can also result from the formation of a heteromeric complex involving donor and recipient gene products.

The set of wheat–barley disomic (Islam *et al.*, 1981) and ditelosomic (Islam, 1983) addition lines was bred from a cross between the wheat cultivar Chinese Spring (CS) and the barley cultivar Betzes (B). A transcriptomic analysis of the six available CS/B disomic chromosome addition lines (the one involving 1H is nonviable) (Cho *et al.*, 2006) has succeeded in allocating the chromosomal origin of 1787 barley transcripts, but this represents only a limited proportion of the full complement of transcripts, estimated to number than 39 000 (Mascher *et al.*, 2017). Here, a line in which the barley 7HL telosome is maintained in a wheat genetic background has been used as a model to characterize interactions between the gene content of the host wheat genome and that of the barley telosome, based on the use of a high capacity transcriptome sequencing platform.

Results

Gene transcription analysis

The RNA-seq procedure yielded on average 14.6 million paired-end mapped reads per sample, of which 98.78 ± 0.81% proved to be assignable to a genomic site (Table S1). Only a small proportion of the reads (2.41% in B, <1% in CS and CS + 7HL) mapped to an unexpected location. The RSEM analysis revealed that 35 301 wheat genes were transcribed by either CS or CS + 7HL at >1 count per million mapped reads, while the equivalent number of 7HL genes located in the genomic window 339–656 Mbp was 923. Of the 35 301 wheat genes, the vast majority (34 341, 97.3%) were not transcribed differentially between CS and CS + 7HL, still leaving 960 which were differentially transcribed. Of these latter genes, 509 were down-regulated (transcript abundance lower in CS + 7HL than in CS) and 451 were up-regulated. The parallel comparison of B *vs* CS + 7HL revealed that of the 923 7HL genes, 389 (42.1%) were differentially transcribed, comprising 233 down-regulated (transcript abundance lower in CS + 7HL than in B) and 156 up-regulated (Table 1). These greatly differing proportions of differential transcription (DT) implied that the effect of a non-native genetic background was much greater than the effect of the presence of an exotic telosome.

Distribution and impact of DT

Four measures were used to quantify the distribution of DT in the wheat genome and along the 7HL chromosome arm; these were (1) the ratio between the number of nondifferentially transcribed ('not-DT') and the total number of transcribed genes [R(not-DT/Trans)], the ratios between the number of up- or down-regulated and the total number of transcribed genes [R(Up/Trans) and R(Down/Trans)] and the mean log fold change (mean log FC). For the 7HL arm, the R(not-DT/Trans) measure indicated that the central segment of the arm was enriched for not-DT genes (Figure 1). The R(Up/Trans) measure was highest at the centromere (0.38) and lowest at the telomere (0.12), while the R(Down/Trans) measure behaved in the opposite fashion (0.15 at the centromere and 0.33 at the telomere). As a result, the mean log FC parameter was at its most positive at the centromere (+0.56) and at its most negative at the telomere (−0.78). Overall, the proximal region of the arm was enriched for up-regulated genes and the distal region for down-regulated ones, leaving the interstitial region neutral. A similar analysis was conducted for the wheat gene complement. On average, a similar number of genes per chromosome were down-regulated (24) as were up-regulated (20), but the numbers were particularly high on chromosome 7A (117 down- and 29 up-regulated genes) (Figure 2). The distribution of DT genes along chromosome 7A is shown in Figure 3. A major feature was the behaviour of all four parameters in two contiguous regions in the distal part of the chromosome's long arm. The first region, a ~10 Mb stretch from 690 to 699 Mbp, harboured 27 transcribed genes, of which 12 were DT (all up-regulated). Along the segment, R(Up/Trans) increased from 0.01 to 0.33, and the mean log FC was +0.52, reflecting a higher average transcript abundance [mean fragments per kilobase of transcript per million mapped read (FPKM): 14.59]. The second region represented the most distal 36 Mbp of the arm (700–736 Mbp), in which there was a set of down-regulated genes, resulting in a R(Down/Trans) value of 0.98, a R(not-DT/Trans) value of just 0.03 and a R(Up/Trans) value of zero. Of the 101 genes present in this region and transcribed in CS, 99 were down-regulated in CS + 7HL and the other two were not-DT genes. The associated mean log FC value was −8.21, and the average transcript abundance in CS + 7HL was extremely low (mean FPKM: 0.25); 25 of the genes recorded an FPKM of zero in every replicate.

Table 1 The differential transcription (DT) in CS + 7HL of both wheat and barley (7HL) genes

	Genes	
	Barley 7HL	Wheat
Transcribed (>1 CPM)	923	35 301
Not-DT	534 (57.85%)	34 341 (97.28%)
DT	389 (42.14%)	960 (2.72%)
Up-regulated	156 (16.90%)	451 (1.28%)
Down-regulated	233 (25.24%)	509 (1.44%)

The terminal region of 7AL in CS + 7HL harbours a 36 Mbp deletion

A comparison of the physical length of flow karyotyped mitotic copies of chromosome 7A suggested that the version harboured by CS + 7HL was approximately 3.0% shorter than the one harboured by CS (Figure 4a). When GAA satellite-based FISHIS patterns of the two versions of chromosome 7A were compared, it was apparent that the most terminal GAA signal on CS 7AL was not present on CS + 7HL 7AL (Figure 4b). The estimated centromeric index (ratios of the long arm to the short arm) of the two versions of 7A revealed that while CS copy is metacentric, CS + 7HL copy was submetacentric. To support the strong indication that the two copies differed as a result of a large

COLOR

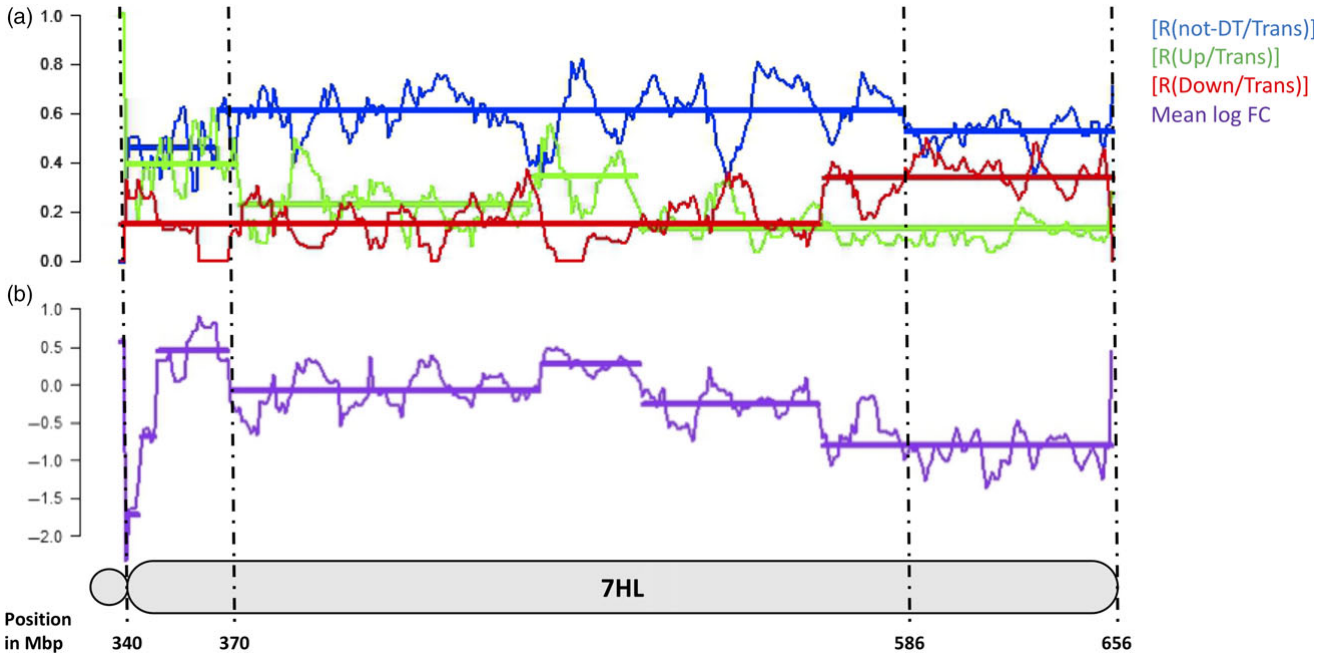


Figure 1 The chromosomal spread of 7HL genes showing altered transcription in CS + 7HL. (a) The ratio of not-DT to transcribed genes [R(not-DT/Trans)], the ratio of up-regulated to transcribed genes [R(Up/Trans)], the ratio of down-regulated to transcribed genes [R(Down/Trans)]. (b) The mean log FC in CS + 7HL and B along the 7HL chromosome. Each segment is represented by a horizontal bar along each measure (change point positions given in Mbp as positions on the 7H pseudomolecule.

COLOR

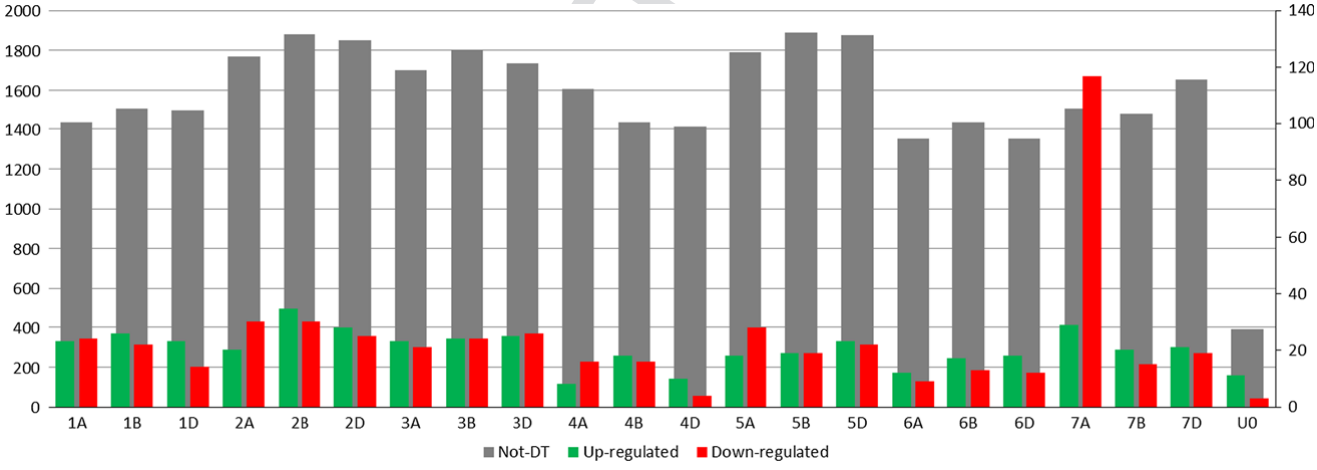


Figure 2 The distribution of not-DT, down- and up-regulated genes across the 21 chromosomes of wheat in CS + 7HL. The number of not-DT genes (grey) is given on the left scale, while down- (red) and up-regulated genes (green) ones are represented on the right scale. Not-DT, up- and down-regulated genes belonging to unanchored scaffolds in the genome assembly have also been represented under the *x*-axis label UO.

deletion event, a set of chromosome 7A molecular markers was developed spanning the 685–736 Mbp region of 7AL (Table S2). When these were used to genotype CS and CS + 7HL, the results confirmed the loss in the latter line of a region between 700 and 736 Mbp of 7A (Figure 4c). The same deletion was also found to be present in each of the six CS/B whole chromosome addition lines, as confirmed by C-banding (Figure S1).

The distal deletion on 7AL does not induce DT of homeologous genes

The regions on chromosome arms 7BL (699–750 Mbp) and 7DL (610–636 Mbp) are homeologous to the 7AL deletion harbour, respectively, 143 and 122 genes. Of these, just six of the 7B genes and two of the 7D ones were transcribed at a higher level in CS + 7HL than in CS (Figure 5a). The R(DT/Trans) values are consistent with the value of ca. 3% which pertains elsewhere in the genome. There was, therefore, no apparent effect of the deletion on the transcription of the 7B and 7D homeologs. The homeologous segment of 7HL (640–656 Mbp) harbours 130 transcribed genes, of which 55 were classified as DT genes in the contrast B *vs* CS + 7HL. As described above (Figure 1), the proportions of DT and not-DT genes in this region did not differ significantly from those obtained in its proximally adjacent segment (600–639 Mbp), which harbours a similar gene density (167 genes, of which 79 were classified as DT genes). Using stringent sequence comparison parameters and collinearity

COLOR

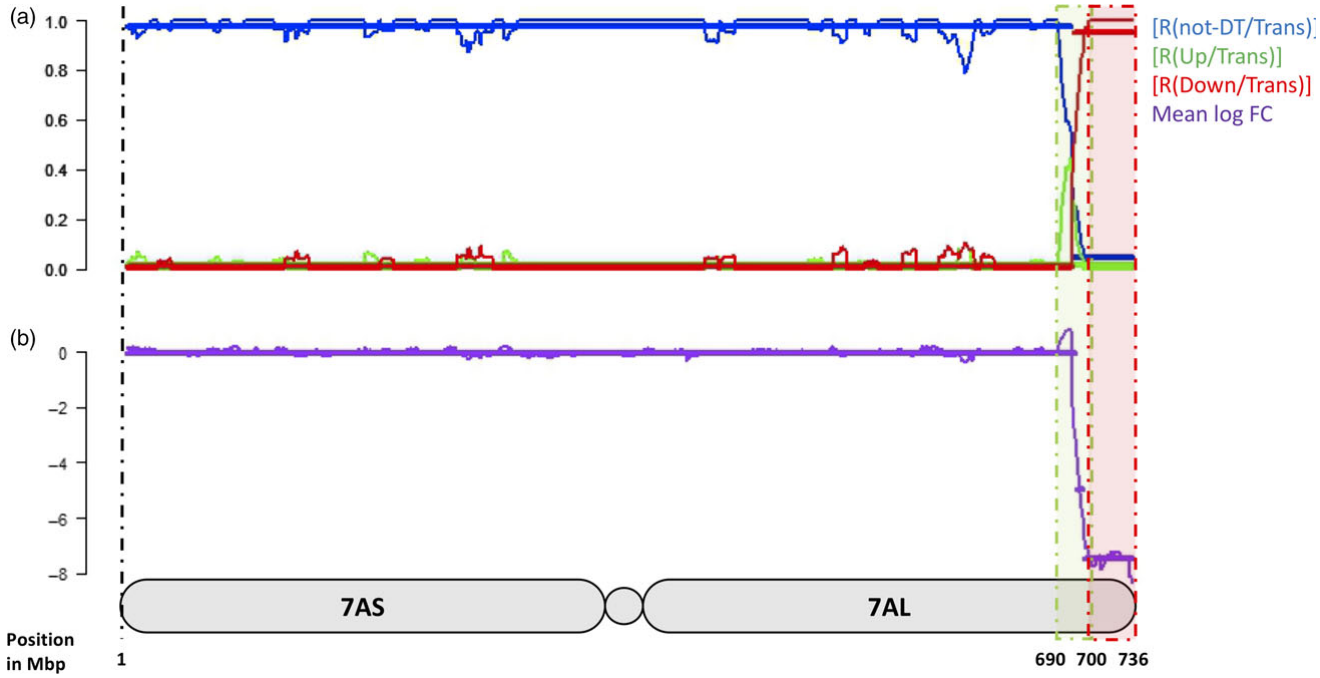


Figure 3 The chromosomal spread of 7A genes showing altered transcription in CS + 7HL. (a) The ratio of not-DT to transcribed genes [R(not-DT/Trans)], the ratio of up-regulated to transcribed genes [R(Up/Trans), the ratio of down-regulated to transcribed genes [R(Down/Trans)]. (b) The mean log FC in CS + 7HL and CS along the 7A chromosome. Each segment is represented by a horizontal bar along each measure (change point positions given in Mbp as positions on the 7A pseudomolecule.

guided strategy, we defined a set of 7HL genes homeologous to 7AL and selected 1 : 1 relationships to investigate whether the loss of 7AL genes induced DT of 7HL syntenic genes in CS + 7HL. The transcription of the homeologous (40%) and nonhomeologous (60%) genes in both the deleted and the nondeleted regions showed similar distributions of homeologs/nonhomeologs of not-DT (50/50, ±0.47), up-regulated (55/45 ± 7.07) or down-regulated (17/83, ±0.7) genes (Figure 5b). The abundance of transcript from the homeologous genes was greater than that from the nonconserved ones, which was the case for both the deleted and the nondeleted regions: the median FPKM for homeologous genes in the deleted region was 10.53 in B and 4.89 in CS + 7HL; for the nonhomeologous genes, the median FPKMs were, respectively, 5.80 and 1.11 (Figure 5c). Overall, therefore, there was no effect of the deletion on the level of transcription of the 7HL gene complement, although a large proportion of the 7HL segment harboured homeologs of the genes lost as a result of the deletion.

The distal deletion on 7AL is partially compensated for by orthologs on 7HL

Of the 923 7HL genes transcribed in B, 702 (76%) were transcribed at a detectable level in CS + 7HL: 375 transcripts were represented by a count of 1–5 FPKM and 327 by >5 FPKM; of the latter, 92 were strongly transcribed (20–1360 FPKM). Around 60% of these genes had at least one transcriptionally active wheat homeolog, so are likely to contribute to the transcript pool. Among the set of highly transcribed 7HL genes, 47 were homeologs of a 7AL gene deleted in CS + 7HL; of these, 26 were at least half as abundantly transcribed in CS + 7HL as was their 7AL homeolog in CS (Figure 6). The indication was, therefore, that the presence of 7HL partially compensated for the loss of transcripts due to the deletion event on 7AL.

Biological relevance of alterations in the pattern of wheat and barley gene transcription

A gene ontology (GO) analysis of the set of wheat DT genes showed that among the 451 up-regulated genes, there was an enrichment for 174 terms for biological processes and molecular functions, relating mostly to the response to biotic, abiotic and oxidative stresses. In particular, the GO terms 'immune system process', 'response to biotic stimulus', 'immune response' and 'ubiquitin-like protein ligase binding' were prominent, as were terms relating to secondary metabolism synthesis such as 'phenylpropanoid metabolic process' and 'secondary metabolite biosynthetic process' (Table S3). Meanwhile, the 410 down-regulated genes strongly featured those relating to energy production, such as 'cellular response to starvation', 'glycolipid biosynthetic process' and 'liposaccharide metabolic process' (Table S4). The implication was that the alterations to the wheat transcriptome induced by the presence of the barley telosome are nonrandom, perhaps reflecting a response to a loss in homoeostasis at the protein and metabolite level. A similar analysis was also applied to the set of genes present in the 7AL deletion, but no significant GO term enrichment was revealed. When the up-regulated set of 156 7HL genes was analysed, there was a significant enrichment for 26 GO terms, notably 'protein synthesis', 'peptide metabolic process', 'translation and amide biosynthetic process' (Table S5). The set of 233 down-regulated 7HL genes were enriched for 14 GO terms, potentially associated in some way with sexual incompatibility: the predominant terms were 'cell recognition', 'single organism reproductive process', 'reproduction', 'multiorganism process' and 'multicellular organismal process' (Table S6). Overall, the GO term enrichment analysis of the DT genes indicated that the altered transcription of recipient and donor genes in CS + 7HL affected a subset of biological functions.

COLOR

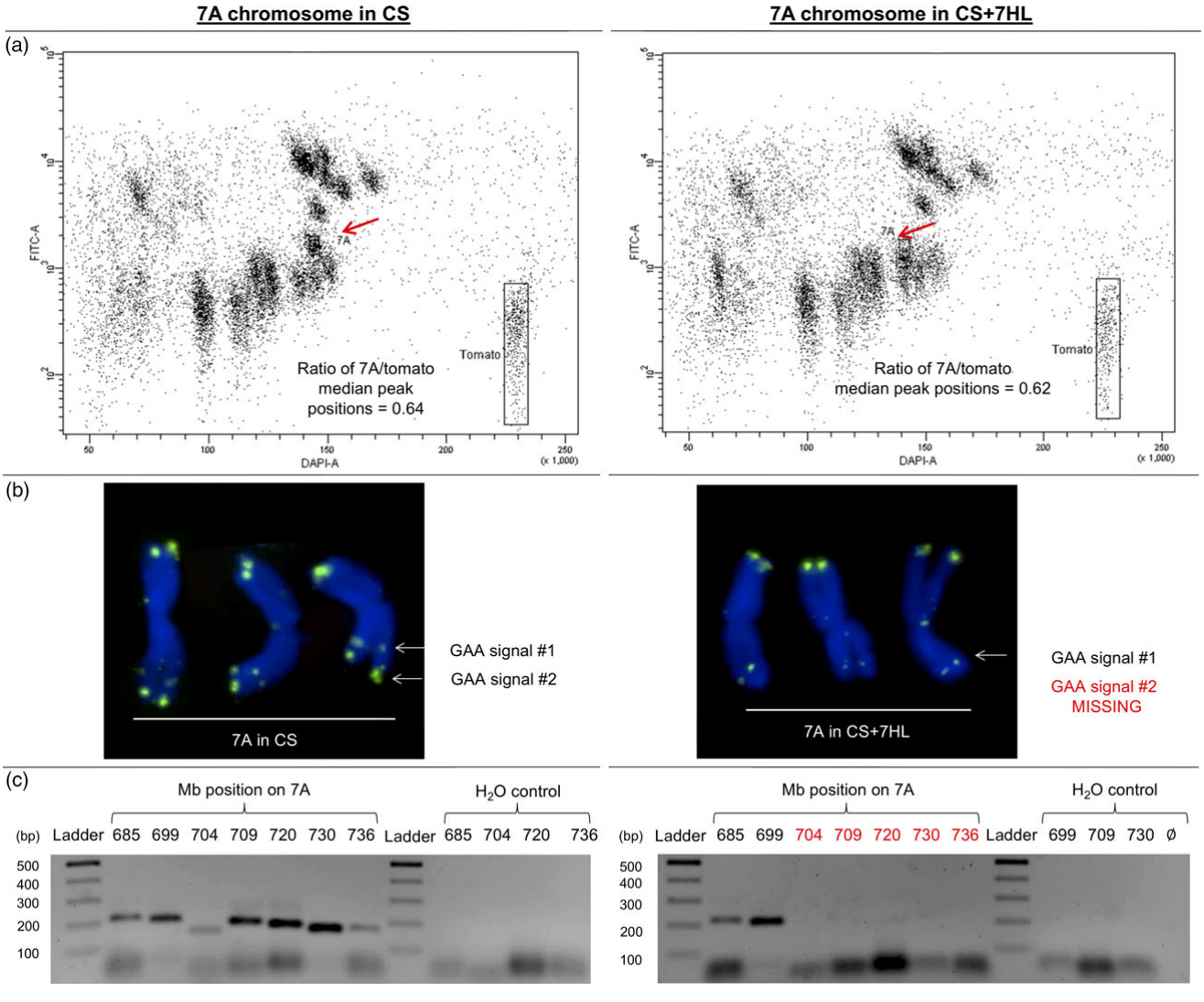


Figure 4 The 7AL arm in the CS/B addition lines carries a major deletion. The deletion is apparent given (a) a size difference in the 7A copies flow-sorted from CS and CS + 7HL, (b) the loss of a GAA FISH site (green) at the distal end of the chromosome and (c) the PCR-based genotyping of 7A deleted region on genomic DNA from CS and CS + 7HL.

Discussion

The RNA-seq platform was used here to capture the transcriptomes of CS, CS + 7HL and B, and applied a more stringent set of criteria that is normally used for mapping reads to a genomic sequence. This was thought necessary to ensure discrimination between homeologs. Rather than aiming to maximize sequencing depth, it was considered a better use of resources to increase the level of biological replication. The approach allowed the transcript abundance of over 35 000 wheat and nearly 1000 barley (7HL) genes to be estimated in the CS + 7HL line, representing, respectively, 31.8% and 34.5% of the annotated gene content. The replication also provided a high measure of confidence to the identification of the 960 wheat and 389 barley genes classified as DT genes in two contrasts CS *vs* CS + 7HL and B *vs* CS + 7HL.

The introgression of 7HL into wheat altered the recipient's and donor's transcriptomes to a varying extent

Wide hybridization is often associated with profound rearrangements of the host and exotic species chromosomal structure, along with alterations to both the DNA sequence and the epigenome; ultimately, these have a profound effect on the transcriptome of the wide hybrid (Buggs *et al.*, 2011; Chen and Ni, 2006; Comai, 2000; Comai *et al.*, 2003; Feldman and Levy, 2005, 2012; Feldman *et al.*, 1997; Gernand *et al.*, 2005; Kashkush *et al.*, 2002; Levy and Feldman, 2004; Pumphrey *et al.*, 2009; Xu *et al.*, 2014). With respect to the CS/B cross, it was not possible to stabilize the initial allohaploid by chromosome doubling, as has been achieved for a range of other wheat-based wide crosses, most notably in the wheat/rye combination triticale. At the time the CS-B hybrid was made (Islam *et al.*, 1981), large-scale transcriptomic analyses were not feasible, and as a result, the impact of genomic shock could not be investigated at this level, although it is known that the early back-cross generations following the production of the CS/B allohaploid were highly abnormal phenotypically (Islam *et al.*, 1981). Based on the experience with stabilizable wide hybrids, it is more than probable that combining two such distantly related species as wheat and barley would have induced major disruptions to the transcriptome, affecting both the barley and wheat components. The present study has revealed, however, that only 3% of wheat genes showed

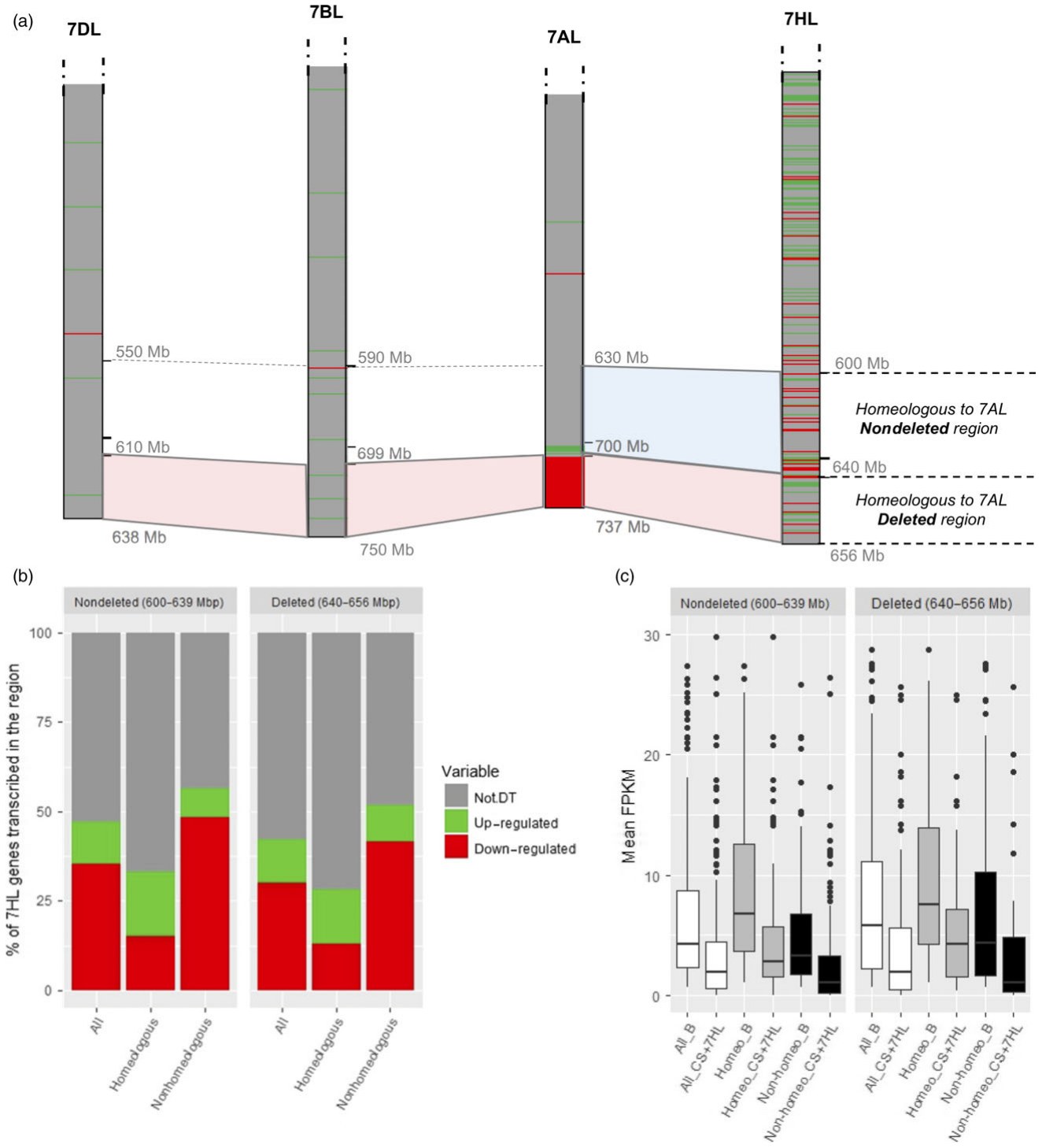
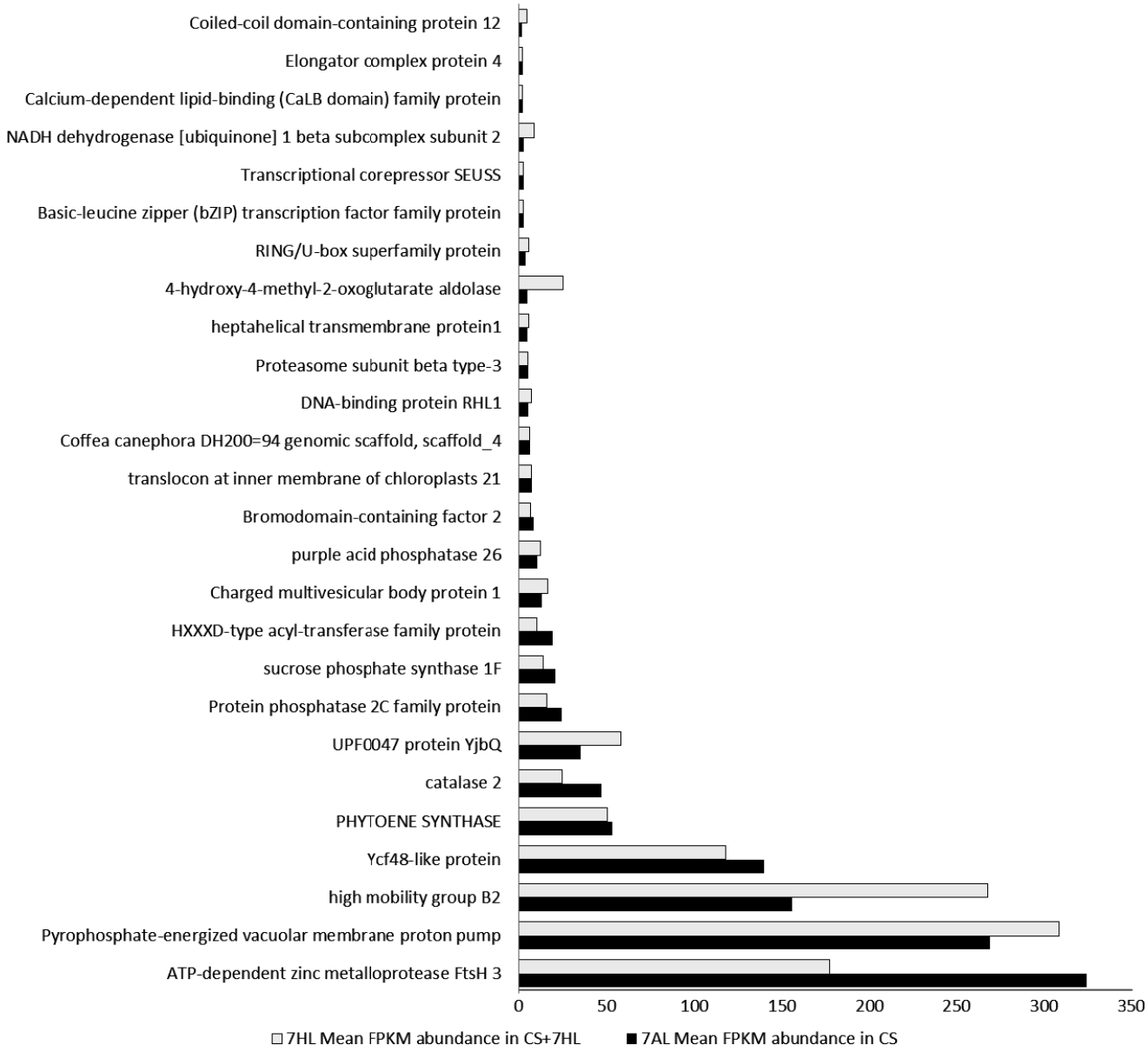


Figure 5 Compensation for genes deleted in the 7A copy present in CS + 7HL by their homeologs in 7B, 7D and 7HL. (a) (Non)-compensation at the level of transcription, (b) the distribution of not-DT, up- and down-regulated genes in the whole (All), homeologous and nonhomeologous portions of 7HL chromosome arm homeologous to the nondeleted (7HL positions 600–639 Mbp) or deleted 7AL region (7HL positions 640–656 Mbp), (c) Boxplot representation of transcript abundance (FPKM) of All, homeologous and nonhomeologous 7HL genes present in CS + 7HL and in B mapping to the regions homeologous to either the nondeleted or deleted regions of 7AL.

any evidence of altered transcription in the CS + 7HL line, while ~42% of the barley genes on 7HL were affected in this way. The relatively mild effect of the presence of 7HL on the wheat transcriptome probably reflects the stabilizing effect of the two generations of back-crossing imposed to reduce the barley complement present in the founding CS/B allohaploid to a single chromosome and to restore that of wheat to its euploid state (Islam, 1983; Islam *et al.*, 1981). In contrast, a large proportion of the genes on 7HL appear to have experienced DT, much of which was likely induced by the genomic shock occurring within the founder allohaploid and its immediate progeny. Although the barley genes affected are distributed along the entire length of the 7HL arm, the ratio of genes experiencing DT proved to be higher at the arm ends: at the proximal end, a majority of the genes affected

LOW RESOLUTION FIG



17 **Figure 6** Partial compensation at the level of transcription for genes deleted in the 7AL copy present in CS + 7HL by their orthologs in 7HL. The partial compensation (at least 50%) of 26 genes deleted in the 7AL chromosome present in CS + 7HL is represented by the abundance given as mean FPKM across all bio-replicated of the 7AL gene copy in CS (dark blue) and the 7HL ortholog in CS + 7HL (clear blue). On *y*-axis is given the functional annotation of the barley 7HL genes according Mascher *et al.* (2017).

were up-regulated, while at the distal end, the majority was down-regulated. A small number (13) of the barley genes have been fully silenced, suggesting a lack of the necessary regulatory sequences; for the rest, their altered transcription intensity implies either the absence of a fully functional regulatory environment and/or an effect of gene dosage due to the presence of their wheat homeologs. Many (~76%) of the barley genes were highly transcribed in the CS + 7HL line, although the comparison with their transcription in B reveals various degrees of regulation in the wheat recipient background. The evidence suggests that most of the genes present on 7HL are regulated by wheat sequences, although for some, this heterologous regulation appears unable to exert the same level of control as managed by native sequences within a barley context. There is no evidence supporting dominance over their transcription exerted by any of the three wheat subgenomes.

The biological relevance and regulation DT genes

The genes which were differentially transcribed were distributed throughout the wheat genome and along the entire length of chromosome 7HL, implying a level of stochasticism. However, when subjected to a GO analysis, there was evidence for nonrandomness, as only a subset of biological functions was affected. Many of the up-regulated genes on barley 7HL were associated with protein translation, synthesis and modification, while on the other hand, the down-regulated ones were related to reproduction and sexual compatibility, encoding a range of transcription factors, hormone receptors, receptor-like protein kinases and certain stress-related proteins. Notably, many of the barley 7HL genes which were up-regulated in the CS + 7HL line encoded products controlling cellular basal functions such as protein synthesis, whereas the down-regulated genes encoded

products involved in reproduction and adaptation; the latter genes may contribute to the reduced fertility of the CS + 7HL line in comparison with the other CS-B ditelosomic addition lines described by Islam (1983). The data indicate that regulatory elements within the wheat genome exert at best partial control over the transcription of the 7HL genes, possibly reflecting the length of the time which has passed since the evolutionary divergence of barley from wheat. The polarized regulation of barley gene transcription (with the transcription of genes at the 7HL telosome's proximal end tending to be favoured) observed in CS + 7HL could arise from the status of telosome's chromatin being unusual or may be a consequence of topology of the telosome within the interphase nucleus (Branco and Pombo, 2007; Elcock and Bridger, 2010; Schubert *et al.*, 2014). The present data set is inadequate to address this issue, as it is limited to a single tissue and developmental stage, while the topology of 7HL in the addition line nucleus is hard to ascertain (Grob *et al.*, 2014; Mukhopadhyay *et al.*, 2013; Schubert *et al.*, 2014). Further transcriptomic, epigenetic and spatial analyses of the host and alien chromosomes in other tissues and developmental stages would be needed to reveal the mechanisms impacting the donor and recipient gene transcription in the CS + 7HL line.

The CS + 7HL addition line harbours a large deletion in 7AL

The CS/B addition lines have been used to identify the genetic basis of a number of traits (Ashida *et al.*, 2007; Bilgic *et al.*, 2007; Kato *et al.*, 2008; McCallum *et al.*, 2004; Sakai *et al.*, 2009; Shi and Endo, 1997; Tang *et al.*, 2011; Yuan *et al.*, 2002). Here, it has been demonstrated that they all share a large (36 Mbp) deletion at the distal end of 7AL. According to the annotation of the 700–736 Mbp region of the 7A pseudomolecule (IWGSC, 2018), the deletion has resulted in a loss of >500 genes. A similarly sized deletion in this region has also been identified among the products induced by the action of a gametocidal chromosome (Endo and Gill, 1996). The fact that the deletion is also present in all six of the CS/B disomic addition lines implies that it could not have arisen at the same time as the formation of the 7HL telosome. Rather, it must have already been present during the development of the addition lines. Islam *et al.* (1981) state that, although 19 independent CS/B allohaploids were produced, only one (which harboured the expected chromosome number of 28) was used for back-crossing to CS to produce the disomic (and later the ditelosomic) addition lines. The most plausible scenario is therefore that the deletion was fixed in the CS stock used as the recurrent back-cross parent to develop the addition lines. At the level of karyotype, the size difference between the wild-type and the deletion-harbouring copy of chromosome 7A is not easy to identify. At the time when the addition lines were created, there was neither any understanding of the ubiquity of deletions in wheat nor had marker technology developed to a sufficient level to allow for such a deletion to have been identified via genotyping. In the meantime, it has been recognized that CS is not a homogeneous genotype (Mott and Wang, 2012). As demonstrated by Kubalakova *et al.* (2002), the bread wheat genome is known to harbour a variety of translocations, deletions and duplications, some of which arise as a result of the genomic shock caused by interspecific hybridization (Bento *et al.*, 2008; Feldman *et al.*, 1997; Fu *et al.*, 2013; Liu *et al.*, 1998). Thus, the performance of individual addition and other introgression lines may be affected not just by the presence of the donor chromosome(s) or chromosome segment(s), but also by hidden variation in the karyotype of the recipient species.

Conclusions

This comparative analysis of transcription in bread wheat, barley and the wheat–barley 7HL addition line has identified that the addition line transcriptome had some unexpected features. The cluster of down-regulated wheat genes on chromosome 7A was shown to be due not to an interaction with the barley telosome, but instead was revealed to be due to a 36 Mbp deletion lying at the distal end of the long arm, likely present in the stock of CS used to generate the disomic addition line set. The genes experiencing DT in the CS + 7HL line were randomly distributed in terms of their genomic location, but appeared to be nonstochastic in relation to their function. This set of genes provides a resource for investigating the molecular basis of DT, which may involve regulatory sequences, epigenetic changes or the organization of chromosome domains in the interphase nucleus. These analyses will be facilitated by the availability of high-quality wheat and barley genome sequence and its associated annotation. The present study has provided a set of genes of potential relevance for determining the gene networks responsible for interspecific incompatibility between wheat and its related species.

Materials and methods

Sample preparation and sequencing

The plant material used in this study comprised CS wheat (2n = 6x = 42, AABBDD), Betzes barley (B; 2n = 2x = 14, HH) and the CS/B addition line harbouring a pair of 7HL telosomes (CS + 7HL; AABBDD + 7HL″). Grain of CS + 7HL and of the six viable whole chromosome CS/B addition lines (+2H through +7H) were obtained from the National BioResource Project, Komugi (shigen.nig.ac.jp/wheat/komugi/strains/nbrpDetailAction.do?strainId=LPGKU2105). The mitotic chromosome number of each aneuploid line was checked to confirm the presence of the expected chromosome complement, following the Kopecky *et al.* (2007) procedure. Plants were raised under a 16 h photoperiod and a day/night temperature regime of 20 °C/16 °C. When they had reached Zadoks stage 11 (first leaf emerged), the oldest leaf from each of three plants of each line was pooled into a single sample (~25 mg), snap-frozen in liquid nitrogen and held at −80 °C. Six biological replicates were collected for each line. RNA was extracted using an RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and treated with RNase-free DNase I (Qiagen) to remove contaminating genomic DNA. The integrity of the RNA was assessed using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA), and an RNA integrity number (RIN) was calculated. Only samples with a RIN >7 were retained for library construction and sequencing, carried out at POPS-IPS2 platform (http://www.ips2.u-psud.fr/spip.php?article213). cDNA libraries were generated for each sample using a TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA). The samples were randomly pooled by six and sequenced on 2 × 100 bp paired-end flow cells using HiSeq2000 device (Illumina), generating ~30 million reads per sample. Ribosomal RNA sequences were removed with the help of sortMeRNA program (Kopylova *et al.*, 2012), reads were cut after the first uncalled base ('N') and low-quality reads (Phred score <Q30) were discarded using Trimmomatic software (Bolger *et al.*, 2014).

RNA-seq data have been deposited in the sequence read archive under the BioProject accession number PRJNA412872.

Transcript analysis

Each of the reads for each of the 18 samples (six replicates of each of CS, B and CS + 7HL) was mapped against the 'High Confidence' coding sequences of both wheat (IWGSC RefSeq v1.0, see IWGSC, 2018) and barley (IBSC RefSeq, see Mascher *et al.*, 2017) using Bowtie v1.0.0 software (Langmead *et al.*, 2009) with no mismatches allowed. The quantification of transcripts abundance was performed using RSEM v1.2.8 software (Li and Dewey, 2011). The assignment of DT was based on two comparisons: one between CS and CS + 7HL, and the other between B and CS + 7HL. Transcripts recording a count of less than one per million mapped reads were ignored, and the counts were then normalized by the effective library size using the trimmed mean of the logarithm of the transcript abundance ratios (TMM); DT was identified using edgeR package version 3.18.1 (Robinson *et al.*, 2010) of the R software version 3.4.1, based on a generalized linear model where the log of the mean expression is a linear function of the factor 'group' defining all biological replicates of either CS, B or CS + 7HL. The false discovery rate for calling DT was set as 5%. Up-/down-regulation refers to genes for which the transcript abundance was significantly higher/lower in CS + 7HL than in either CS or B, without applying fold change threshold to see all changes. Finally, the TMM-normalized counts were converted into FPKMs to enable comparisons between different genes within or between samples.

Genomic distribution of the genes responsible for the transcripts

The density of the transcribed genes was calculated on the basis of a sliding 10 Mbp window, with a step of 1 Mbp along each chromosome. Similarly, the ratios of down- and up-regulated to transcribed transcripts [R(Down/Trans), R(Up/Trans)] were calculated. The mean logarithm of the fold change in transcript abundance (logFC) in the comparisons CS *vs* CS + 7HL and B *vs* CS + 7HL was determined for the same genomic windows. Each of these measures was processed using the R package changepoint v2.2.2 program (Killick and Eckley, 2014) employing the 'Pruned Exact Linear Time' method (Killick *et al.*, 2012).

Flow cytometry

The procedure used for preparing metaphase chromosome suspensions and subsequent FISHIS labelling has been described by Vrana *et al.* (2016). A crude suspension of tomato nuclei obtained from chopped leaf was added to the chromosome suspensions to serve as an internal reference for the estimation of the DNA content of chromosome 7A in both CS and CS + 7HL. Flow cytometry and sorting were conducted using a FACSAria flow cytometer (BD Biosciences, San José, CA) equipped with blue laser (488 nm, 100 mW) for FITC fluorescence excitation and a UV laser (355 nm, 100 mW) for DAPI fluorescence excitation. For the purpose of PCR, 500 mitotic copies of chromosome 7A from each of the two lines were sorted into a tube containing 10 μL deionized water.

Cytogenetic analysis

Chromosomes 7A sorted from CS and CS + 7HL were subjected to FISH, using as probes either a GAA microsatellite or an Afa family repeat, to identify individual chromosomes. The FISH protocol followed that of Vrana *et al.* (2016), except that in case of metaphase spreads, the denaturation period was extended to 2 min at 80 °C. C-banding analysis was performed following the protocol previously described by (Endo, 2011).

Genotypic analysis

Genes present on the group 7 chromosomes for which there were known to be a full set of three homeologs were identified using the homeologs data set provided with the IWGSC RefSeq_v1.0 (IWGSC, 2018). The genomic sequence of each homeolog was processed using GSP software (Wang *et al.*, 2016) to allow PCR primers to be designed with a Tm of 50–65 °C and a GC content of 50%–60%; the predicted amplicon size range was limited to the range 150–250 bp. The selected primer sequences were scanned against both the IWGSC RefSeq (IWGSC, 2018) and IBSC RefSeq (Mascher *et al.*, 2017) sequences to remove any which were likely nonspecific for the 7A homeolog. Finally, PCRs were run on a set of templates (genomic DNA from CS, CS + 7HL, along with sorted 7A chromosomes from CS and CS + 7HL).

Definition of homeologous and nonhomeologous gene sets

Sets of homeologous and nonhomeologous genes between 7HL and wheat subgenomes were determined by whole genome comparison of protein sequences using BLASTp (Altschul *et al.*, 1990) leveraged by MCScanX software (Wang *et al.*, 2012) with *E*-value threshold of 10e-5 and a size of synteny block of 10 genes. Finally, only 1 : 1 relationships were retained for each subgenome comparison.

Gene ontology term enrichment analysis

Gene ontology terms were associated with each of the wheat and barley genes according to functional annotations provided by, respectively, IWGSC (2018) and Mascher *et al.* (2017). Ontologizer2.0 software (Bauer *et al.*, 2008) was applied to identify GO term enrichment. The full set of transcribed genes for wheat and 7HL genes was used as reference comparison set for the enrichment analysis of the genes deleted in 7AL and for the up- and down-regulated genes in CS + 7HL. *P*-values were evaluated using the Parent–Child–Union method (Grossmann *et al.*, 2007) and then adjusted by applying the Benjamini–Hochberg correction (Benjamini and Hochberg, 1995). GO terms considered as enriched were retained on the basis of adjusted *P*-value thresholds of 0.01 (wheat) and 0.05 (7HL).

Acknowledgements

The authors thank Dr. Etienne Paux for fruitful scientific discussions; Petr Navrátil for outstanding informatics support; Radim Svačina, Helena Tvardíková, Marie Vyhlídalová for excellent technical support. 'National BioResource Project, Komugi' is acknowledged for maintaining and providing the plant material. This research was financially backed by the Czech Science Foundation (award P501/12/G090) and by the Ministry of Education, Youth and Sports of the Czech Republic (award LO1204 from the National Program of Sustainability I). The computing infrastructure required was provided by the National Grid Infrastructure MetaCentrum (grant No. LM2010005 under the program 'Projects of Large Infrastructure for Research, Development, and Innovations').

Conflict of interest

The authors declare no conflict of interest.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Ashida, T., Nasuda, S., Sato, K. and Endo, T.R. (2007) Dissection of barley chromosome 5H in common wheat. *Genes Genet. Syst.* **82**, 123–133.

Bauer, S., Grossmann, S., Vingron, M. and Robinson, P.N. (2008) Ontologizer 2.0 – a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics*, **24**, 1650–1651.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B (Methodological)*, **57**, 289–300.

Bento, M., Pereira, H.S., Rocheta, M., Gustafson, P., Viegas, W. and Silva, M. (2008) Polyploidization as a retraction force in plant genome evolution: sequence rearrangements in triticale. *PLoS ONE*, **3**, e1402.

Bilgic, H., Cho, S., Garvin, D.F. and Muehlbauer, G.J. (2007) Mapping barley genes to chromosome arms by transcript profiling of wheat–barley ditelosomic chromosome addition lines. *Genome*, **50**, 898–906.

Blakeslee, A. (1937) Dédoublement du nombre de chromosomes chez les plantes par traitement chimique. *C. R. Acad. Sci. Paris*, **205**, 476–479.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Börner, A., Ogbonnaya, F.C., Röder, M.S., Rasheed, A., Periyannan, S. and Lagudah, E.S. (2015) *Aegilops tauschii* introgressions in wheat. In *Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and Genomics* (Molnár-Láng, M., Ceoloni, C. and Doležel, J., eds), pp. 245–271. Cham: Springer International Publishing.

Branco, M.R. and Pombo, A. (2007) Chromosome organization: new facts, new models. *Trends Cell Biol.* **17**, 127–134.

Brozynska, M., Furtado, A. and Henry, R.J. (2016) Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnol. J.* **14**, 1070–1085.

Buggs, R.J., Zhang, L., Miles, N., Tate, J.A., Gao, L., Wei, W., Schnable, P.S. *et al.* (2011) Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Curr. Biol.* **21**, 551–556.

Chen, Z.J. and Ni, Z. (2006) Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *BioEssays*, **28**, 240–252.

Chen, P., Liu, W., Yuan, J., Wang, X., Zhou, B., Wang, S., Zhang, S. *et al.* (2005) Development and characterization of wheat- *Leymus racemosus* translocation lines with resistance to Fusarium Head Blight. *Theor. Appl. Genet.* **111**, 941–948.

Cho, S., Garvin, D.F. and Muehlbauer, G.J. (2006) Transcriptome analysis and physical mapping of barley genes in wheat–barley chromosome addition lines. *Genetics*, **172**, 1277–1285.

Comai, L. (2000) Genetic and epigenetic interactions in allopolyploid plants. *Plant Mol. Biol.* **43**, 387–399.

Comai, L., Madlung, A., Josefsson, C. and Tyagi, A. (2003) Do the different parental 'heteromes' cause genomic shock in newly formed allopolyploids?. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358**, 1149–1155.

Curtis, T. and Halford, N.G. (2014) Food security: the challenge of increasing wheat yield and the importance of not compromising food safety. *Ann. Appl. Biol.* **164**, 354–372.

Elcock, L.S. and Bridger, J.M. (2010) Exploring the relationship between interphase gene positioning, transcriptional regulation and the nuclear matrix. *Biochem. Soc. Trans.* **38**, 263–267.

Endo, T.R. (1988) Induction of chromosomal structural changes by a chromosome of *Aegilops cylindrica* L. in common wheat. *J. Hered.* **79**, 366–370.

Endo, T.R. (2007) The gametocidal chromosome as a tool for chromosome manipulation in wheat. *Chromosome Res.* **15**, 67–75.

Endo, T.R. (2011) Cytological dissection of the triticeae chromosomes by the gametocidal system. *Methods Mol. Biol.* **701**, 247–257.

Endo, T. and Gill, B. (1996) The deletion stocks of common wheat. *J. Hered.* **87**, 295–307.

Fatih, A.M. (1983) Analysis of the breeding potential of wheat-Agropyron and wheat-Elymus derivatives. I. Agronomic and quality characteristics. *Hereditas*, **98**, 287–295.

Fedak, G. (1999) Molecular aids for integration of alien chromatin through wide crosses. *Genome*, **42**, 584–591.

Feldman, M. and Levy, A.A. (2005) Allopolyploidy – a shaping force in the evolution of wheat genomes. *Cytogenet. Genome Res.* **109**, 250–258.

Feldman, M. and Levy, A.A. (2012) Genome evolution due to allopolyploidization in wheat. *Genetics*, **192**, 763–774.

Feldman, M., Liu, B., Segal, G., Abbo, S., Levy, A.A. and Vega, J.M. (1997) Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. *Genetics*, **147**, 1381–1387.

Feuillet, C., Langridge, P. and Waugh, R. (2008) Cereal breeding takes a walk on the wild side. *Trends Genet.* **24**, 24–32.

Friebe, B., Jiang, J., Raupp, W.J., McIntosh, R.A. and Gill, B.S. (1996) Characterization of wheat-alien translocations conferring resistance to diseases and pests: current status. *Euphytica*, **91**, 59–87.

Fu, S., Yang, M., Fei, Y., Tan, F., Ren, Z., Yan, B., Zhang, H. *et al.* (2013) Alterations and abnormal mitosis of wheat chromosomes induced by wheat-rye monosomic addition lines. *PLoS ONE*, **8**, e70483.

Gernand, D., Rutten, T., Varshney, A., Rubtsova, M., Prodanovic, S., Bruss, C., Kumlehn, J. *et al.* (2005) Uniparental chromosome elimination at mitosis and interphase in wheat and pearl millet crosses involves micronucleus formation, progressive heterochromatinization, and DNA fragmentation. *Plant Cell*, **17**, 2431–2438.

Griffiths, S., Sharp, R., Foote, T.N., Bertin, I., Wanous, M., Reader, S., Colas, I. *et al.* (2006) Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. *Nature*, **439**, 749–752.

Grob, S., Schmid, M.W. and Grossniklaus, U. (2014) Hi-C analysis in Arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of *Drosophila*. *Mol. Cell*, **55**, 678–693.

Grossmann, S., Bauer, S., Robinson, P.N. and Vingron, M. (2007) Improved detection of overrepresentation of Gene-Ontology annotations with parent–child analysis. *Bioinformatics*, **23**, 3024–3031.

Gupta, P., Varshney, R.K., Sharma, P. and Ramesh, B. (1999) Molecular markers and their applications in wheat breeding. *Plant Breeding*, **118**, 369–390.

Hurni, S., Brunner, S., Stirnweis, D., Herren, G., Peditto, D., McIntosh, R.A. and Keller, B. (2014) The powdery mildew resistance gene Pm8 derived from rye is suppressed by its wheat ortholog Pm3. *Plant J.* **79**, 904–913.

Islam, A.K.M.R. (1983) Ditelosomic additions of barley chromosomes to wheat. In *Proceedings of the Sixth International Wheat Genetics Symposium* (Sakamoto, S., ed.), pp. ???–???. ????: ????.

Islam, A., Shepherd, K.W. and Sparrow, D.H.B. (1981) Isolation and characterization of euplasmic wheat–barley chromosome addition lines. *Heredity*, **46**, 161–174.

IWGSC (2018) Shifting the limits in wheat research and breeding through a fully annotated and anchored reference genome sequence. *Science*, submitted.

Jiang, J., Friebe, B. and Gill, B. (1993) Recent advances in alien gene transfer in wheat. *Euphytica*, **73**, 199–212.

Kashkush, K., Feldman, M. and Levy, A.A. (2002) Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics*, **160**, 1651–1659.

Kato, K., Kidou, S. and Miura, H. (2008) Molecular cloning and mapping of casein kinase 2 alpha and beta subunit genes in barley. *Genome*, **51**, 208–215.

Kerber, E.R. and Aung, T. (1999) Leaf rust resistance gene lr34 associated with nonsuppression of stem rust resistance in the wheat cultivar canthatch. *Phytopathology*, **89**, 518–521.

Kerber, E.R. and Dyck, P.L. (1973) Inheritance of stem rust resistance transferred from diploid wheat (*Triticum monococcum*) to tetraploid and hexaploid wheat and chromosome location of the gene involved. *Can. J. Genet. Cytol.* **15**, 397–409.

Killick, R. and Eckley, I.A. (2014) changepoint: an R package for changepoint analysis. *J. Stat. Softw.* **1**, ???–???.

Killick, R., Fearnhead, P. and Eckley, I.A. (2012) Optimal detection of changepoints with a linear computational cost. *J. Am. Stat. Assoc.* **107**, 1590–1598.

Kopecky, D., Allen, D.C., Duchoslav, M., Dolezel, J. and Lukaszewski, A.J. (2007) Condensation of rye chromatin in somatic interphase nuclei of Ph1 and ph1b wheat. *Cytogenet. Genome Res.* **119**, 263–267.

Kopylova, E., Noe, L. and Touzet, H. (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, **28**, 3211–3217.

Kruse, A. (1973) Hordeum × Triticum hybrids. *Hereditas*, **73**, 157–161.

Kubalakova, M., Vrana, J., Cihalikova, J., Simkova, H. and Dolezel, J. (2002) Flow karyotyping and chromosome sorting in bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **104**, 1362–1372.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.

Levy, A.A. and Feldman, M. (2004) Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. *Biol. J. Lin. Soc.* **82**, 607–613.

Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.

Liu, B., Vega, J.M. and Feldman, M. (1998) Rapid genomic changes in newly synthesized amphiploids of Triticum and Aegilops. II. Changes in low-copy coding DNA sequences. *Genome*, **41**, 535–542.

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T., Radchuk, V. *et al.* (2017) A chromosome conformation capture ordered sequence of the barley genome. *Nature*, **544**, 427–433.

McCallum, B.D., Tekauz, A. and Gilbert, J. (2004) Reaction of a diverse collection of barley lines to Fusarium head blight. *Plant Dis.* **88**, 167–174.

Mondal, S., Rutkoski, J.E., Velu, G., Singh, P.K., Crespo-Herrera, L.A., Guzman, C., Bhavani, S. *et al.* (2016) Harnessing diversity in wheat to enhance grain yield, climate resilience, disease and insect pest resistance and nutrition through conventional and modern breeding approaches. *Front. Plant Sci.* **7**, 991.

Mott, I.W. and Wang, R.R.C. (2012) Genetic variation among laboratory accessions of Chinese Spring wheat (*Triticum aestivum* L.). *Plant Genet. Resour.* **10**, 97–100.

Mujeeb-Kazi, A. (1995) *Intergeneric crosses: hybrid production and utilization*. *CIMMYT Research Report (CIMMYT)*.

Mukhopadhyay, P., Singla-Pareek, S.L., Reddy, M.K. and Sopory, S.K. (2013) Stress-mediated alterations in chromatin architecture correlate with down-regulation of a gene encoding 60S rpL32 in rice. *Plant Cell Physiol.* **54**, 528–540.

Murashige, T. and Skoog, F. (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473–497.

Pumphrey, M., Bai, J., Laudencia-Chingcuanco, D., Anderson, O. and Gill, B.S. (2009) Nonadditive expression of homoeologous genes is established upon polyploidization in hexaploid wheat. *Genetics*, **181**, 1147–1157.

Rayburn, A. and Gill, B. (1985) Use of biotin-labeled probes to map specific DNA sequences on wheat chromosomes. *J. Hered.* **76**, 78–81.

Rey, E., Molnár, I. and Doležel, J. (2015) Genomics of wild relatives and alien introgressions. In *Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and Genomics* (Molnár-Láng, M., Ceoloni, C. and Doležel, J., eds), pp. 347–381. Cham: Springer International Publishing.

Reynolds, M.P., Calderini, D.F., Condon, A.G. and Rajaram, S. (2001) Physiological basis of yield gains in wheat associated with the LR19 translocation from *Agropyron elongatum*. *Euphytica*, **119**, 139–144.

Riley, R. and Chapman, V. (1958) Genetic control of the cytologically diploid behaviour of hexaploid wheat. *Nature*, **182**, 713–715.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Sakai, K., Nasuda, S., Sato, K. and Endo, T.R. (2009) Dissection of barley chromosome 3H in common wheat and a comparison of 3H physical and genetic maps. *Genes Genet. Syst.* **84**, 25–34.

Schubert, V., Rudnik, R. and Schubert, I. (2014) Chromatin associations in Arabidopsis interphase nuclei. *Front. Genet.* **5**, 389.

Shi, F. and Endo, T.R. (1997) Production of wheat–barley disomic addition lines possessing an *Aegilops cylindrica* gametocidal chromosome. *Genes Genet. Syst.* **72**, 243–248.

Stirnweis, D., Milani, S.D., Brunner, S., Herren, G., Buchmann, G., Peditto, D., Jordan, T. *et al.* (2014) Suppression among alleles encoding nucleotide-binding-leucine-rich repeat resistance proteins interferes with resistance in F1 hybrid and allele-pyramided wheat plants. *Plant J.* **79**, 893–903.

Tang, J., Ohyama, K., Kawaura, K., Hashinokuchi, H., Kamiya, Y., Suzuki, M., Muranaka, T. *et al.* (2011) A new insight into application for barley chromosome addition lines of common wheat: achievement of stigmasterol accumulation. *Plant Physiol.* **157**, 1555–1567.

Vrana, J., Capal, P., Cihalikova, J., Kubalakova, M. and Dolezel, J. (2016) Flow sorting plant chromosomes. *Methods Mol. Biol.* **1429**, 119–134.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H. *et al.* (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49.

Wang, Y., Tiwari, V.K., Rawat, N., Gill, B.S., Huo, N., You, F.M., Coleman-Derr, D. *et al.* (2016) GSP: a web-based platform for designing genome-specific primers in polyploids. *Bioinformatics*, **32**, 2382–2383.

Wulff, B.B.H. and Moscou, M.J. (2014) Strategies for transferring resistance into wheat: from wide crosses to GM cassettes. *Front. Plant Sci.* **5**, 692.

Xu, C., Bai, Y., Lin, X., Zhao, N., Hu, L., Gong, Z., Wendel, J.F. *et al.* (2014) Genome-wide disruption of gene expression in allopolyploids but not hybrids of rice subspecies. *Mol. Biol. Evol.* **31**, 1066–1076.

Yamamoto, M. and Mukai, Y. (1989) Application of fluorescence in situ hybridization to molecular cytogenetics of wheat. *Wheat Information Service*, **69**, 30–32.

Yuan, Y., Chen, X., Xiao, S., Islam, A. and Xin, Z. (2002) Identification of wheat–barley 2H alien substitution lines. *Acta Botanica Sinica*, **45**, 1096–1102.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1 All CS/B addition lines carry a major deletion in the 7AL chromosome arm. The same deletion as detected in the CS + 7HL ditelosomic addition line is present in the six CS/B whole chromosome addition lines, as shown by C-banding. Chromosomes 7A are indicated on the pictures by red arrows.

Table S1 The RNA-seq mapping fitness.

Table S2 Molecular markers used to genotype 7AL deletion in CS and CS + 7HL.

Table S3 The gene ontology terms enriched in the set of wheat up-regulated genes in CS + 7HL.

Table S4 The gene ontology terms enriched in the set of wheat down-regulated genes in CS + 7HL.

Table S5 The gene ontology terms enriched in the set of barley 7HL genes up-regulated genes in CS + 7HL.

Table S6 The gene ontology terms enriched in the set of barley down-regulated genes in CS + 7HL.

6.3 Appendix III

Transcriptomic analysis of interactions between host and alien genomes in wheat-barley interspecific introgression

Elodie Rey, Gabriel Keeble-Gagnère, Sandrine Balzergue, Ludivine Soubigou-Taconnat, Véronique Brunaud, Marie-Laure Martin-Magniette, Takashi R. Endo, Michael Abrouk, Jan Bartoš, Rudi Appels and Jaroslav Doležel

In abstracts of: "International Conference on Genomics and Bioinformatics". Izmir, Turkey, 2016.

TRANSCRIPTOMIC ANALYSIS OF INTERACTIONS BETWEEN HOST AND ALIEN GENOMES IN WHEAT-BARLEY INTERSPECIFIC INTROGRESSION

Elodie Rey[1], Gabriel Keeble-Gagnère[2], Sandrine Balzergue[3], Ludivine Soubigou-Taconnat[3], Véronique Brunaud[4], Marie-Laure Martin-Magniette[4], Takashi R. Endo[5], Michael Abrouk[1], Jan Bartoš[1], Rudi Appels[2] and Jaroslav Doležel[1]

(1) Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic
(2) Murdoch University, 90 South Street, Perth, WA 6150, Australia
(3) POPS Transcriptomic Platform, IPS2, CNRS, INRA, Univ 12 Paris Sud, Univ Evry, Univ Paris-Diderot, Sorbonne Paris-Cite, Univ Paris-Saclay, 91405 Orsay, France
(4) Bioinformatics & Predictive Genomics team, IPS2, CNRS, INRA, Univ 12 Paris Sud, Univ Evry, Univ Paris-Diderot, Sorbonne Paris-Cite, Univ Paris-Saclay, 91405 Orsay, France
(5) Department of Plant Life Science, Faculty of Agriculture, Ryukoku University, Shiga 520-2194, Japan
Corresponding author: Elodie Rey – Institute of Experimental Botany rey@ueb.cas.cz

Background

For wheat as for many other crops, alien introgression - the process during which genes/gene compexes are transferred between species via interspecific hybridization - is an attractive approach to recover genetic diversity that was lost during domestication and breeding, and to introgress other valuable traits for agronomical improvement of cultivated species. In order to increase the efficiency of this process, we need to understand the molecular mechanisms controlling the function of alien genes and their interaction with the host. We have chosen wheat-barley ditelosomic addition line 7HL, which was obtained after hybridization between bread wheat cv. Chinese Spring and barley cv. Betzes, as a model system to investigate those mechanisms.

Aim

The aim of our project is to provide novel information on the interaction between a host genome and introgressed chromosomes and their parts, both at genome and transcriptome levels. In the present analysis, we have sequenced mRNA of both wheat and barley parents and the wheat-barley chromosome-arm addition line 7HL to reveal changes in the patterns of expression resulting from the intergenomic interactions.

Experiment

18 RNA-seq samples
3 genotypes:
Wheat cv Chinese spring (CS)
Barley cv Betzes
Chromosome add. line CS^{+7HL}
6 biological replicates

Total RNA extraction
Leaves from plantlets state Zadok 1.1
Qiagen RNeasy Plant Mini Kit

mRNA selection + cDNA library construction
Illumina TruSeq® RNA Sample Prep. v2

Sequencing
Illumina HiSeq 2000 - PE (2 x 100 bp)
3 pools of 6 samples per lane
~ 30M reads/sample

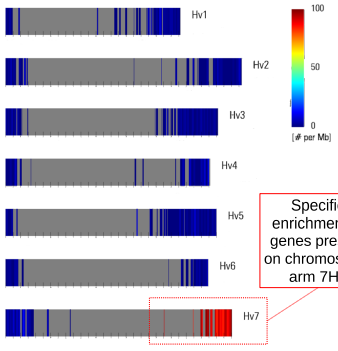
RNA-seq reads mapping
Against genome assemblies: wheat CS (Ensembl) + barley Morex (MIPS)
+ CDS annotations (.gtf)
STAR (v. 2.5.0c)

→ **Transcripts assembly**
Cufflinks
Merging assemblies
From all replicates of all genotypes
Cuffmerge

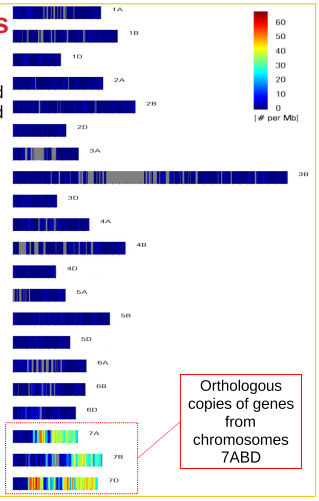
→ **Gene expression analysis**
RSEM (bowtie1)
+
edgeR (GLM analysis)

Verification of the transcripts assembly

Comparative analysis between 7HL transcripts and Genomes of barley (left) and wheat (right) and displayed by chromoWIZ:



Specific enrichment of genes present on chromosome arm 7HL



Orthologous copies of genes from chromosomes 7ABD

Challenge

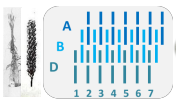
- Wheat and barley are relatives, and thus share a high degree of similarity between genic sequences.
- Potential problem to accurately estimate transcripts abundance of highly conserved genes between the two genomes in CS^{+7HL}.
- A BLASTn analysis between barley (60,760 CDS) and wheat (88,144 CDS) allows us to identify those problematic genes

results parsed following criteria of % identity (ID) and alignment length (AL)

Criteria	Comparison	Nb. of relationships
ID 90% / AL 100bp	Barley *vs* Wheat	~52 000
	7H *vs* Wheat	~7 600
	7H *vs* 7ABD	~5 500
ID 90% / complete AL	**7H *vs* Wheat**	**~400**
	7H *vs* 7ABD	**~360**

Results

Wheat
Triticum aestivum cv. Chinese spring
2*n*=6*x*=42 (AADDBB)



55,165 transcripts

11.6 MYA divergence

2195 (2,8%)
problematic transcripts
(not-DE between wheat and barley)

Barley
Hordeum vulgare cv. Betzes
2*n*=2*x*=14 (HH)

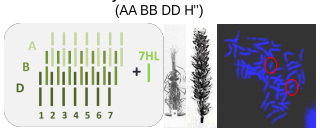
21,411 transcripts
1740 from chr. 7HL

98,5% of wheat-host transcripts are **not-DE** in presence of barley 7HL alien genes

98,7% of transcripts expressed by chr. 7HL are **down-regulated** in CS^{+7HL}

wheat-barley ditelosomic add. line 7HL
(AA BB DD H")

1,4% is **DE** between CS^{+7HL} and wheat-parent:
513 transcripts are **down-regulated** in CS^{+7HL} and 292 are **up-regulated**



1,14% is **not-DE** compared to Betzes parent

CONCLUSIONS AND PROSPECTS

The RNA-seq analysis of wheat-host and barley-alien genes expression revealed a **global down-regulation of barley 7HL introgressed genes** in the wheat genetic background and a **soft impact** of the introgression **on the expression of the wheat host genes.** We are currently investigating DE genes within their genomic context: gene structure, position along the chromosome, orthologous relationships with the host/alien genome, functional annotation and gene ontology enrichment. This information will allow to (1) infer the effect of structural and functional features on the changes in gene expression observed, which might provide clues to understand (2) how the changes in gene expression are directed, and (3) how do they impact on the introgression line phenotype.

This work was supported by the Czech Science Foundation (award no. P501/12/G090) and the National Program of Sustainability I (grant award LO1204).

6.4 Appendix IV

Transcriptomic analysis of barley 7HL genes introgressed in wheat

Elodie Rey, Sandrine Balzergue, Ludivine Soubigou-Taconnat, Véronique Brunaud, Endo R. Takashi, Michael Abrouk, Jan Bartos and Jaroslav Dolezel

In abstracts of: "European Wheat Aneuploid Conference (EWAC) 2015 International Conference". Lublin, Poland, 2015.

Transcriptomic analysis of barley 7HL genes introgressed in wheat:

Towards the understanding of interaction between a host genome and introgressed chromosomes

Elodie Rey[1], Sandrine Balzergue[2], Ludivine Soubigou-Taconnat[2], Véronique Brunaud[3], Endo R. Takashi[1,4], Michael Abrouk[1], Jan Bartos[1] and Jaroslav Dolezel[1]

(1) Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic
(2) The Transcriptomic Platform, Unité de recherche en génomique végétale URGV-INRA Evry, France
(3) Bioinformatics & Predictive Genomics team, Unité de recherche en génomique végétale URGV-INRA Evry, France
(4) Department of Plant Life Science, Faculty of Agriculture, Ryukoku University, Shiga 520-2194, Japan

Corresponding author: Elodie Rey - Institute of Experimental Botany - Centre of the Region Haná for Biotechnological and Agricultural Research - Olomouc - Czech Republic - rey@ueb.cas.cz

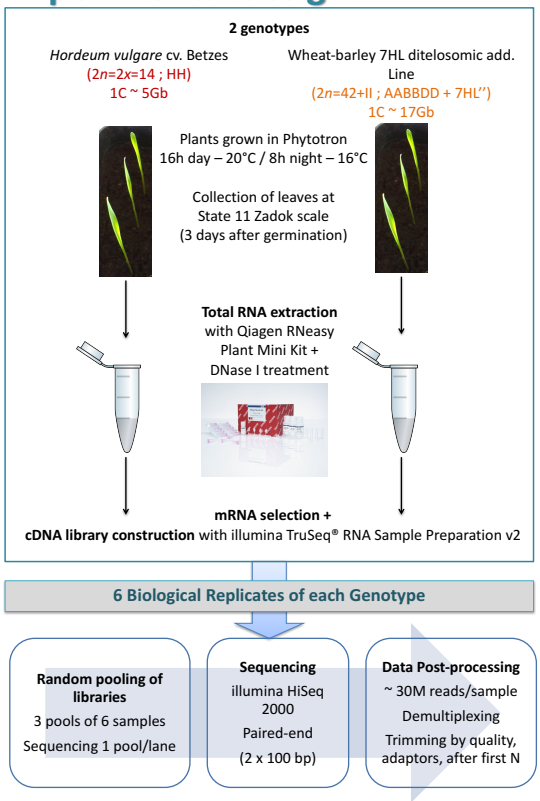
Background

Since more than a century, wheat has been crossed with its wild relatives in attempt to introduce favorable alleles, genes and gene complexes providing resistance to biotic and abiotic stresses and other beneficial traits. Although hybridization with wild species represents a promising approach to improve cultivated varieties of bread wheat, the knowledge of underlying biological mechanisms remains limited and only a few wheat-alien introgression lines have made their way into the agriculture. We have chosen wheat-barley ditelosomic addition line 7HL, which was obtained after hybridization between bread wheat cv. Chinese Spring and barley cv. Betzes, as a model system to investigate those mechanisms.

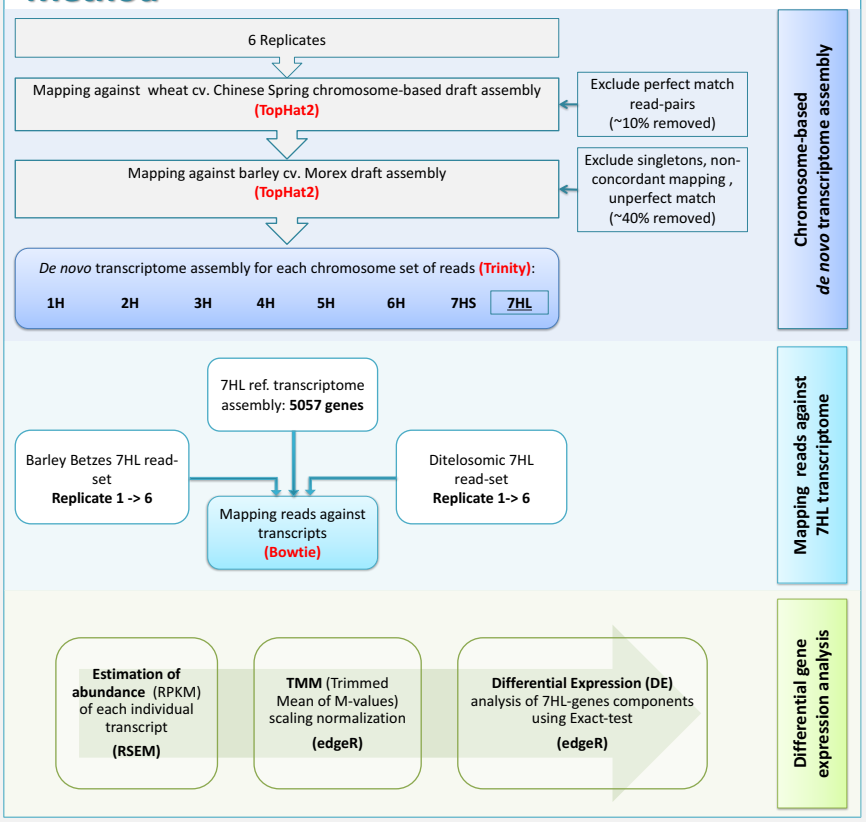
Aim

The aim of our project is to provide novel information on the interaction between a host genome and introgressed chromosomes and their parts, both at genome and transcriptome levels. In the present analysis, we compared expression of 7HL genes in the natural background (barley parent) with the expression of the same gene set in the wheat host background.

Experimental design



Method



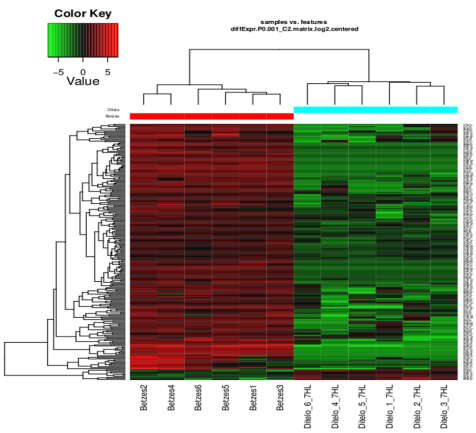
Results

Among the 5057 genes in the 7HL transcript's reference assembly:

- 238 (~4.7%) are differentially expressed (4-fold) between the 2 genotypes
- The majority of those genes are lower expressed in the ditelosomic line
- Several clusters of genes with common expression profiles are visible
- Among them, a small group of genes (~5.5%) show higher expression in the wheat host genome compared to original barley genomic background

Prospects

- Our future work will be focused on improving the statistical analysis in order to better account for biological variation between the replicates of same genotype.
- We want to investigate the mechanisms underlying changes in gene expression observed between the parent and the introgression line. Among others, we would like to study the structure of the genes, their position along the chromosome-arm and the epigenetic state of the chromatin.
- At long term, we plan to apply this method to study expression of wheat-host genes in the presence of the barley-alien genomic segment



Heatmap showing the relative expression level of each gene (rows) in each sample (column).

References

Feuillet, C., Langridge, P., and Waugh, R. (2008). Cereal breeding takes a walk on the wild side. Trends Genet. TIG 24, 24–32.

Islam, A.K.M.R., Shepherd, K.W., and Sparrow, D.H.B. (1981). Isolation and characterization of euplasmic wheat-barley chromosome addition lines. Heredity 46, 161–174.

Jeffrey Chen, Z., and Ni, Z. (2006). Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. BioEssays News Rev. Mol. Cell. Dev. Biol. 28, 240–252.

Acknowledgements

This work was supported by the grant LO1204 from the Czech National Program of Sustainability I.

6.5 Appendix V

Genomics of alien gene transfer in wheat

Elodie Rey, Sandrine Balzergue, Ludivine Soubigou-Taconnat, Véronique Brunaud, Endo R. Takashi, Jan Bartos and Jaroslav Dolezel

In abstracts of: "EUCARPIA Cereals Section – I T M I Joint Conference".

Wernigerode, Germany, 2014.

Genomics of Alien Gene Transfer in Wheat

Elodie Rey[1], Sandrine Balzergue[2], Ludivine Soubigou-Taconnat[2], Véronique Brunaud[3], Endo R. Takashi[1,4], Michaël Abrouk[1], Jan Bartos[1] and Jaroslav Dolezel[1]

(1) Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic
(2) The Transcriptomic Platform, Unité de recherche en génomique végétale URGV-INRA Evry, France
(3) Bioinformatics & Predictive Genomics group, Unité de recherche en génomique végétale URGV-INRA Evry, France
(4) Graduate School of Agriculture, Kyoto University, Japan

Corresponding author: Elodie Rey / Institute of Experimental Botany - Centre of the Region Haná for Biotechnological and Agricultural Research / Olomouc - Czech Republic /rey@ueb.cas.cz

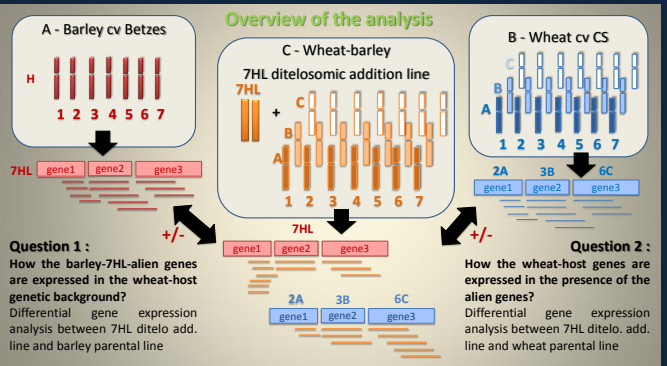
Background

Introducing genes from wild and close relatives into a host genome via interspecific hybridization is a powerful strategy to enhance elite crop gene pools. Its application is especially important in wheat, which feeds ~40% of the world's population and provides ~20% of the total food calories and proteins in human nutrition. It is necessary that productivity and quality of cultivated wheat continue increasing to satisfy the need for food by the growing human population. Thus, pre-breeding programs involve crosses between wheat and its cultivated and wild relatives (Feuillet et al., 2008). Although alien gene introgression has been used increasingly, the knowledge of underlying biological mechanisms remains limited, and little is known about the regulation of alien gene expression after introgression into a different genetic background.

Aim

The study aims to provide novel information to support the use of chromosome-mediated gene transfer in wheat improvement. To do so we assess gene expression in wheat cv. Chinese Spring (CS), barley cv. Betzes and wheat-barley 7HL ditelosomic addition line (Islam et al., 1981). Barley was chosen as a model because it is diploid, a distant relative of wheat, and should induce strong effects in the wheat background. The chromosome arm 7HL carries important genes, including resistance to pathogens, salt tolerance, good dietary compounds and earliness. Comparison of gene expressions in the ditelosomic addition line and the parents should provide novel information to understand the regulation of alien barley genes in a host genome, and reciprocally the impact of alien genes on the expression of the wheat-host genes.

Overview of the analysis



Question 1 :
How the barley-7HL-alien genes are expressed in the wheat-host genetic background?
Differential gene expression analysis between 7HL ditelo add. line and barley parental line

Question 2 :
How the wheat-host genes are expressed in the presence of the alien genes?
Differential gene expression analysis between 7HL ditelo. add. line and wheat parental line

Question 3 : What are the mechanisms under the changes in gene expression observed between the parents and the alien introgression line?
- What is the gene ontology of the differentially expressed genes?
- How does the gene copy number variation affect the expression of duplicated genes?
- How does the genome structure impact on gene expression?

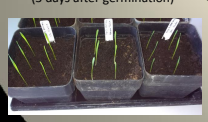
Plant material

3 Genotypes (A, B, C)
A. *Hordeum vulgare* cv. Betzes (2n=2x=14)
B. *Triticum aestivum* cv. Chinese spring (2n=6x=42 ; AABBDD)
C. Wheat-barley 7HL ditelosomic add. Line (2n=42+II ; AABBDD + 7HL'')

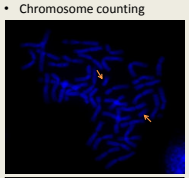
Germination
(3 days – 25°C)

Growing plants in Phytotron
(controlled conditions)
16h day at 20°C
/ 8h night at 16°C

Collecting plant material from leaves for RNA extraction state 11 of the Zadoks scale (3 days after germination)



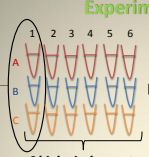
Checking of each 7HL ditelo add. line karyotype
- On seminal roots collected for each 7HL germinated seed
- Cell synchronization by cold treatment + fixation
- Chromosome counting



Karyotype of one wheat-barley 7HL ditelosomic addition line observed on fluorescent microscope after DAPI staining. 42 wheat chromosomes and 2 telosomes are visible.

Experiment design

1 Experiment:
- 3 genotypes (A, B, C)
- 1 RNA sample/genotype
- Each sample = pool of 3 plants

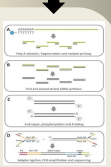


6 biological repeats

Plant material
- Collected from leaves
- Immediately deep frozen in liquid nitrogen
- Kept at -80°C until extraction

Total RNA extraction with Qiagen RNeasy Plant Mini Kit + DNase I treatment

Those steps have been realised in collaboration with the Transcriptomic platform and the Bioinformatics & Predictive Genomics group from URGV-INRA Evry (France):



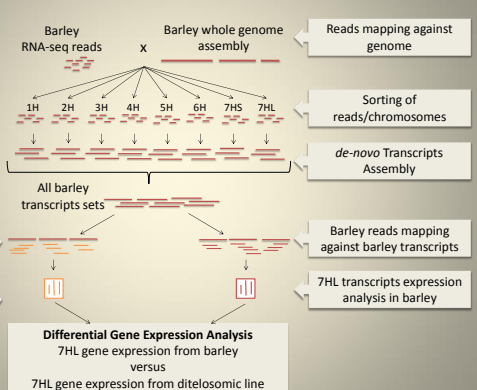
mRNA selection + Library construction with illumina TruSeq® RNA Sample Preparation v2
- Libraries randomly pooled
- 3 pools of 6 samples
- Sequencing of 1 pool/lane
- 3 lanes of sequencing

Sequencing on illumina HiSeq 2000
- Paired-end (2 x 100 bp)

Data Post-processing
- ~ 30M reads/sample
- Demultiplexing
- Trimming by quality, adaptors, after first N

Barley transcriptome analysis workflow

We adopted a genome guided *de-novo* transcripts assembly strategy for the reconstruction of wheat and barley transcriptomes. The aim was to make the best use of the available ressources, and to reduce the complexity of the transcriptome analysis.



The same genome guided *de-novo* transcripts assembly will be applied to wheat.

Discussion

Interspecific hybridization is known to induce a genomic shock (Jeffrey Chen, Z., and Ni, Z., 2006). It affects gene regulation and results in novel gene activation, gene repression, or even gene silencing. In order to analyze the effect of these processes during chromosome-mediated gene transfer, we use wheat-barley telosome addition line as experimental system. Since wheat and barley are distant relatives, we expect those changes to be significant. We also expect that due to the two generations of back-crossing of the F1-amphiploid with wheat, followed by several generations of self-pollination of the ditelosomic line, the changes in gene expressions are stabilized.

Future work

The next step will be to analyze differences in gene expression between the parental wheat and the ditelosomic addition lines. We will implement genome-guided *de-novo* transcripts assembly strategy to discriminate transcripts coming from different copies of genes of the three wheat sub-genomes. Differential expression analysis from both barley and wheat genes should permit identification of co-regulated groups of genes. Gene expression analysis will be complemented by the analysis of epigenetic medications and by identification of structural genome changes induced by the introduction of alien chromosome.

References

Feuillet, C., Langridge, P., and Waugh, R. (2008). Cereal breeding takes a walk on the wild side. Trends Genet. TIG 24, 24–32.

Islam, A.K.M.R., Shepherd, K.W., and Sparrow, D.H.B. (1981). Isolation and characterization of euplasmic wheat-barley chromosome addition lines. Heredity 46, 161–174.

Jeffrey Chen, Z., and Ni, Z. (2006). Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. BioEssays News Rev. Mol. Cell. Dev. Biol. 28, 240–252.

Acknowledgements

This work was supported by the grant LO1204 from the Czech National Program of Sustainability I.

6.6 Appendix VI

EMBO short-term fellowship

Award ASTF 485 – 2015

EMBO

European Molecular Biology Organization

Date: 22 March 2016

Certificate

Elodie Rey

This is to acknowledge that the above named Applicant is a recipient of an EMBO Short Term Fellowship, commencing on 13 November 2015

Place of Fellowship: The University of Melbourne

Melbourne
Australia

Project Title: Interaction between a host plant genome and introgressed chromatin

Duration: 90 days **ASTF No:** 485 - 2015

Prof. Maria Leptin
Director, EMBO

EMBO

Postfach 1022.40
D-69012 Heidelberg
Germany

Meyerhofstraße 1
D-69117 Heidelberg
Germany

Tel: + 49-6221-8891-130 E-mail: fellowships@embo.org Telefax: +49-6221-8891215
Director: Prof. Maria Leptin