library@chuka.ac.ke; www.chuka.ac.ke

DYNAMIC QUANTITATIVE TRAIT LOCI AND COPY NUMBER VARIATION: THE MISSING HERITABILITY OF COMPLEX AGRONOMIC TRAITS

Muraya, M.M.
Department of Plant Sciences, Chuka University, P. O. Box 109-60400, Chuka
Email: moses.muraya@chuka.ac.ke, moses.muraya@gmail.com

Citation
Muraya, M.M.(2016) Dynamic quantitative trait loci and copy number variation: the missing heritability of complex agronomic traits. In: Isutsa, D.K. and Githae, E.W. Proceedings of the Second Chuka University International Research Conference held in Chuka University, Chuka, Kenya from 28th to 30th October, 2015.122-130pp.

ABSTRACT
Genetic studies have identified thousands of loci controlling various agronomic traits, revealing important biological pathways and providing valuable insights into genetic basis of trait variation. However, genome-wide association studies (GWAS) have explained relatively small heritability of most complex traits, leading to the question of 'missing' heritability of complex traits. This study examined the 'missing' heritability and offered clues on the underlying genetic architecture of complex traits, such as biomass accumulation in maize. Twelve main effect and 6-pair of epistasis quantitative trait loci **(**QTL), displaying different patterns of expression at different developmental time points in 261 maize genotypes were used. Copy number variation (CNV) and presence absence variation (PAV) was used to study the genetic architecture in 34 maize genotypes. The identified QTL and CNV were mapped on maize B73 reference genome. A total of 182 genes were found harboured in the detected QTL regions. A complex CNV architecture, such as smaller CNV nested within larger CNV or overlapping CNV regions was detected throughout the maize genome, which may explain the extraordinary traits variation observed in maize. The complex CNV genetic architecture may partly explain the missing heritability. The differential gene expression and their interactions at different developmental time points may also explain the missing heritability. Consequently, the genetic model from final trait values cannot reflect the real gene action during the entire growth and development of a plant. It is necessary to understand the CNV and the dynamics of gene expression for complex traits at different developmental stages as a basis for quantitative trait manipulation.
Keywords: Complex trait; Gene action; Genome-wide association studies

INTRODUCTION
The major goal of plant geneticists is to understand how genetic variation contributes to phenotypic variation in the population. To this end, the genetic sources of phenotypic variation have been a major focus in plant breeding studies aimed at identifying the causes of trait variation, improving agriculture and understanding adaptive processes. Many agronomic traits are complex and controlled by many genes, each with a small additive effect (Bernardo 2008; Zuo and Li J 2014). Genome-wide association studies (GWAs) holds great promise for the dissection of complex traits (Yu and Buckler 2006; Stich and Melchinger 2010). The approach (GWAs) provides a high-resolution method for mapping QTL (to the gene level) based on linkage disequilibrium (Yu and Buckler 2006). Many genetic variants contributing to complex traits have been identified (Bian et al., 2013; Busemeyer et al., 2013; Liu et al., 2014; Würschum et al., 2014; Bullucci et al., 2015; Bac-Molenaar et al., 2015), but when several genes have been linked to a trait, both individual and cumulative effects are small and not enough to explain estimated heritability.

In maize, QTL analysis has been strongly supported through sequencing and assembly of the reference genome (Schnable et al., 2009) and derived genotyping approaches (Ganal et al., 2011) and has been applied to a wide variety of morphological and physiological traits (Hao et al., 2011; Zheng and Liu 2013). However, they explain only a few% of the phenotypic diversity, hence the question 'where is the missing heritability?' (Manolio et al., 2009; Eichler et al., 2010). These and the vast majority of other QTL studies in plants assess the expression of traits at a certain stage, frequently at final harvest (Buckler et al., 2009). Therefore, very limited information has been reported on dynamically acting genetic factors in plants assessed via monitoring trait expression at multiple time points (Bian et al 2013; Busemeyer et al., 2013; Liu et al., 2014; Würschum et al., 2014; Bullucci et al., 2015; Bac-Molenaar et al., 2015). Moreover, structural variation has been recognised as a major contributor to genomic diversity in various organisms (Henrichsen et al., 2009; Diskin et al., 2009; Springer et al., 2009; Conrad et al., 2010; Belo' et al., 2010; Yu et al., 2011)). Maize genomes are rich in structural diversity, including copy number variation (CNV) and presence absence variation (PAV), but these type of variation is still poorly understood (Springer et al., 2009; Belo' et al., 2010; Swanson-Wagner et al., 2010). Recent reports have suggested a role of CNV, either individually or in aggregate, as the cause of hitherto unexplained genetic variation (Springer et al., 2009; Diskin et al., 2009).

The inability to find some genes is due to the fact that rare variants are detectable only when sample size is adequate at the local level (Manolio et al., 2009; Eichler et al., 2010; Luo et al., 2011). In many populations allelic heterogeneity of same gene exists and these are associated with different phenotypes (Bergelson and Roux 2010; Wood et al., 2011; Zhang et al., 2011). The single-marker linkage is also affected by genetic heterogeneity, when multiple major loci are involved and in linkage disequilibrium (LD) with each other (Platt et al., 2010). The epistatic interactions variations normally go undetected because epistasis can only be determined by sequential genome-wide scan of major loci (Storey et al., 2005). The epigenetic variation is a likely source of missing heritability (Johannes et al., 2009). This paper advances two additional possible answers to missing heritability, i.e., the complex CNV/PAV genome architecture and dynamically acting genetic factors in plants. The information on these two genetic variants is limited or lacking. The paper also tries to highlight some of the constraint that maybe faced in an attempt to incorporate of CNV/PAV and dynamic genetic factors (dynamic QTL) in GWAs.

MATERIALS AND METHODS

Copy number variation (CNV) and presence absence of variation (PAV)

A 2.1 M oligonucleotide NimbleGen microarray designed by Roche NimbleGen (Springer et al., 2009) was used in this experiment. Array comparative genomic hybridisation (aCGH) was conducted according to NimbleGen aCGH analysis protocol, using 34 maize inbred lines and B73 as reference genome. Genomic DNA (gDNA) was isolated from leaf tissue of 2-weeks old maize seedling (10 plants per inbred line) using a modified CTAB protocol (Mace et al., 2003). Equal amounts of DNA were pooled from the ten individuals per inbred line to constitute the working gDNA samples. Test (inbred lines) and reference (B73) gDNAs (1µg each) were labelled separately with fluorescent dyes, Cy5 and Cy3 respectively, using NimbleGen dual-colour labelling kit. Labelled gDNAs were then combined and hybridized to the microarrays for 72 hours in a NimbleGen hybridization station at 42° C. The hybridised arrays were scanned at 2µm resolution with an Agilent scanner and images were extracted and analysed with NimbleScan v2.6 software. For each test inbred line, two genome-wide aCGH profiles were obtained, representing the log2-ratios of measured fluorescent intensities for test inbred line vs. B73. All aCGH profiles were normalised and were then analysed by a three-state Hidden Markov Model (HMM) described in Seifert et al (2012) to identify CNV and/or PAV between a test inbred line and the B73 reference genome. Sigmalmap was used to map CNV and/or PAV to their genomic locations.

Dynamically acting genetic factors (Dynamic quantitative trait loci)

Phenotyping

A panel of 261 inbred lines was phenotyped using an automated high-throughput phenotyping system (LemnaTec scanalyzer 3D) for their biomass accumulation from 8 – 42 days after sowing (DAS) in three seasons (2011-2012). The experiment was laid out in an incomplete randomised block design and replicated twelve times. Plants were imaged every day from 8 - 42 DAS. The estimates of fresh shoot biomass were extracted from the digital images taken daily during the growth period (8 - 42 DAS). Then the Integrated Analysis Platform (Klukas et al., 2014) was used to calculate plant biomass volume from images acquired daily as estimates of biomass accumulation during the plant growth period.

Genotyping

The 261 maize inbred lines were genotyped using the Illumina SNP chip MaizeSNP50 with 56,110 evenly spaced SNPs distributes across the ten maize chromosomes (Ganal et al., 2011). A total of 35,682 loci were used after filtering for quality control, which exclude SNPs with rates of missing values above 5%, rates of heterozygotes above 5%, and allele frequencies smaller than 0.05 or larger than 0.95.

Association mapping

A standard linear mixed model based on the BLUEs of the 261 maize lines estimated across the three seasons for eleven time points (11, 22, 26, 28, 30, 32, 34, 36, 38 and 42 DAS) independently was used to perform genome-wide association mapping scans (Yu et al., 2006). The marker effects were assumed

fixed marker and genotype as random effects. The population structure was corrected using the kinship matrix (*Jiang* et al., 2014). A two-dimensional genome scan based on markers with significant main effects was performed to study marker-marker interactions. The model included the detected main effect QTL as co-factors as well as the main and interaction effects of the marker pair under consideration (Würschum et al., 2011). Significance of marker-trait associations was tested based on the Wald F statistic. The Bonferroni-Holm procedure (Holm 1979) was used to detect markers with significant ($P < 0.05$) main and interaction effects. The detected SNPs were then mapped on maize B73 reference genome (B73 version v1 release 4a.53) and genes in linkage disequilibrium with these SNPs identified.

RESULTS

Complexity of maize genome

This study reveals structural genomic variation dispersed along the maize chromosomes, which includes thousands of CNV/PAV. On average 18,737 CNV/PAV were detected between any randomly selected pair of inbred lines. Maize has a very complex genome architecture. The detected CNV exhibited genomic architectural complexity in form of smaller CNV within larger ones and CNV with inter-lineage variation in extent of displaying different start and end points (Figure 1). The CNV found in multiple inbred lines displayed inter-lineage variation, with frequently different breakpoints.

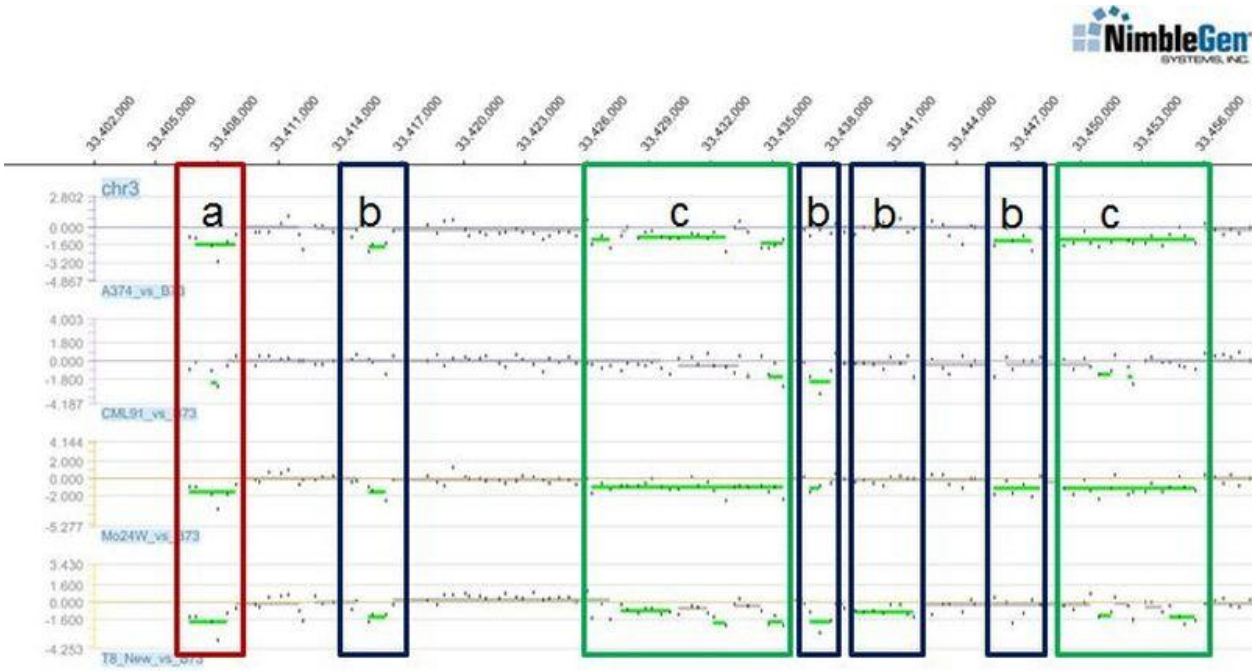

Figure 1: Copy number variation (CNV; 16.6kb) showing different copy number variation (CNV) characteristics in chromosome 3 in four maize inbred maize: (a) CNV displaying CNV in the four genotypes albeit in different copy numbers, (b) CNV indicating that one or more of the inbred line lack a CNV at this genomic position, and (c) CNV exhibiting a complex genomic architectural CNV in between the inbred lines, probably suggesting different inter-lineage CNV breakpoint or existence of small CNV within larger one.

Association mapping

Association mapping scan revealed that dynamic QTL for biomass accumulation were well distribute throughout the maize genome, being detected in nine of the ten maize chromosomes (Figure 2). The dynamic QTL analysis showed that different loci with major effects are expressed at different

developmental time point (Figure 2). Epistasis mapping scanning also revealed that different loci interact at different developmental time points (Figure 3). The results of the study imply that there is upregulation and downregulation of genes controlling complex traits (e.g. biomass accumulation) at different growth and developmental stages of the plant.

A total of 182 genes were found to be harboured in the detected QTL regions, of which 54 have been annotated (Table S1). Two of the genes, AC215286.2_FG002 and GRMZM5G859954, are categorised as cold response genes. The GRMZM5G859954, a main effect locus, located at the bottom of chromosome 2, is expressed at early stages of seedling development (11 DAS; Figure 2 and 3). The gene AC215286.2_FG002 at the top of chromosome 1 interact with other genes at the mid of chromosome 9, at the early stages of seedling development (11 DAS). Cold response genes would be important at early stages of seedling development as they are likely to influence early biomass production during the late spring. A couple of genes are involved in transport and photosynthesis.

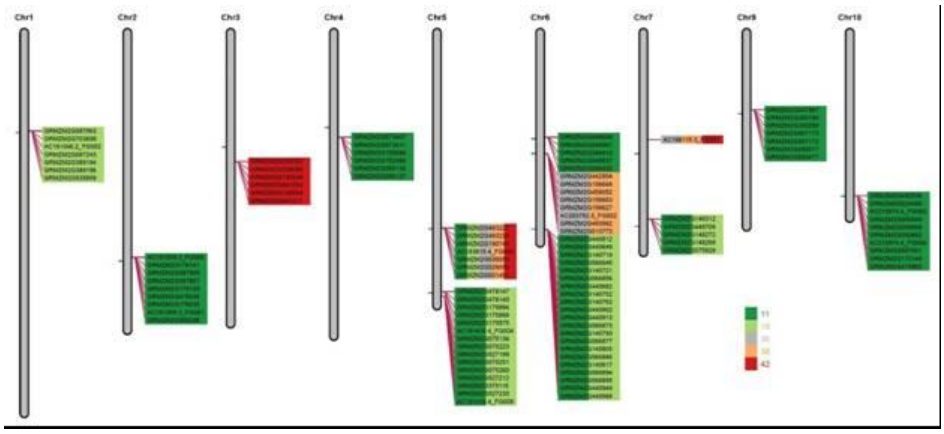

Figure 2: The figure displays significant 12 SNP associations (Holm-Bonferroni =0.05) for maize biomass accumulation and production at different growth time point and genes harbouring the SNP or within 55 kb up- and down-stream of the SNP

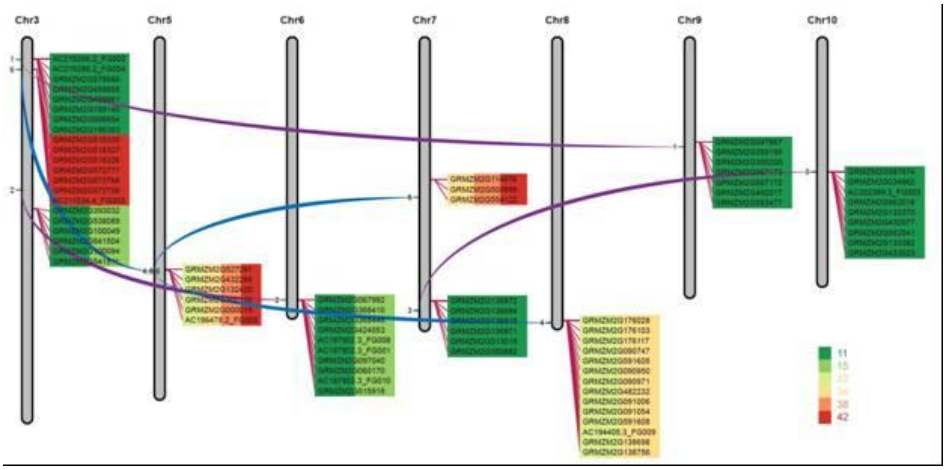

Figure 3: The figure displays 6 pairs of significant interactions (Holm-Bonferroni = 0.05) effects among different loci for maize biomass accumulation and production at different growth time point and genes harbouring the SNP or within 55 kb up- and down-stream of the SNP. Purple and blue connections indicate single and multiple QTL interaction, respectively

DISCUSSION

Understanding the heritability of complex traits requires a more comprehensive assessment of plant genetic variation. Though GWAs have been used to dissect many complex traits, majority of these studies uses SNPs-phenotype associations (Bian et al., 2013; Busemeyer et al., 2013; Liu et al., 2014; Würschum et al., 2014; Bullucci et al., 2015; Bac-Molenaar et al., 2015). Consequently, limiting dissection of trait variation to SNPs genome variation. This study showed that maize genome is populated with structural variants, CNV/PAV. The findings are consistent with findings from other studies which have shown that plant genomes are rich in structural diversity (Springer et al., 2009; Belo' et al., 2010; Yu et al., 2011). Yet, this type of genetic variants has not been accounted for in genome-wide association mapping. Structural variation includes inversions, translocations and CNV. Copy number variation (CNV) describes DNA sequences (usually considered to be larger than 1 kb in size) that are present in genomes being compared albeit in different numbers of copies (Springer et al., 2009). The most extreme form of CNV is PAV, which refers to sequences that are present in some genomes but missing in others (Springer et al., 2009; Swanson-Wagner et al., 2010).

These genomic imbalances (CNV/PAV) represent a special class of genetic variants that can potentially affect many genes and pathways in a single individual. Though SNPs are more frequent, CNV affect larger genomic sequences and thus have the potential to elicit stronger effects, including changing gene structure and dosage, altering gene regulation and exposing recessive alleles (Henrichsen et al., 2009; Zhang et al., 2009). In this regard, CNV can be considered as a major source of genetic variation, thus potentially contributing to genetic diversity and evolution and consequently contributing to the missing heritability. Gene duplication serves as an evolutionary mechanism for functional innovation (Zhang 2003). Gene turnover in the form of rapid expansion or contraction of gene families has been put forward as a possible explanation of phenotypic divergence (Zhu et al., 2007; Perry et al., 2008). In human, available data suggest that CNV genes are highly variable among individuals, and enriched genes are associated with environmental interaction (Alkan et al., 2009). In human, CNV has been used to explain missing heritability in disorders such as schizophrenia and autism (Stefansson et al 2008; The International Schizophrenia Consortium, 2008). This study demonstrated that even among inbred lines, in which the genetics is simplified to a comparison between two genomes (test inbred line and B73), there is variation in genomic architecture among lines, leading to complex phenotypes. The differences in genomic architecture reflect the complex, often opposing effects of selection, population history, migration and mutation rates. These structural variants can account for a large portion of genetic variation among individual genotypes and therefore could account for some of missing heritability.

A large number of genes acting and interacting at a different plant developmental time were detected in this study for the complex trait, biomass accumulation. This suggested that the expression of a complex trait is a result of action of many genes that may behave differentially during the entire growth and developmental time of a given individual plant, and that gene expression is modified by the interactions genes at different growth time points. Simple phenotypes such as susceptibility to disease is due to genetic variants of large effect (Min-Oo et al*.,* 2003; Diez et al., 2003), but complex phenotypes (e.g., variation in lipoproteins) have complex genetic architecture due to the joint action of very many loci of small effect (Valdar et al., 2006). The estimation of the positions and effects of QTL is of central importance for marker assisted selection (Zheng and Liu 2013). In underground networks, most genes work together with close related genes, and it is possible that the effects of one gene on heritability cannot be found without knowing the effects of the others. In complex trait variation also exist in the extent to which epistasis shapes a phenotype. Epistasis implies one gene can mask the effect of another or several genes can work together. For example, two genes acting at given time may each add one gram to the biomass on their own, but together or even acting at different growth time point they could add five grams. This study demonstrates that the genetic model from final biomass cannot reflect the real gene action during the entire development of the plant. It is therefore, necessary to understand the dynamics of gene

expression for biomass accumulation as a trait at different developmental stages as a basis for quantitative trait manipulation.

It is worth noting that, though the genome sequence information and excellent genomic tools are in place for major crop species (Schnable et al., 2009; Paterson et al., 2009; Ganal et al., 2011), phenotyping remain the major bottleneck in systematic quantification of phenotypes. Genome-wide association studies (GWAs) for dynamically acting genetic factors are constraint by phenotyping. Conventional phenotyping procedures are generally labour-intensive, time consuming, lower throughput, costly, and frequently destructive to plants (e.g., fresh or dry weight determination). Moreover, measurements are often taken at certain times or at particular developmental stages, leading to a phenotyping bottleneck (Furbank and Tester, 2011). The current high-throughput genotyping platforms are amicable to automation, non-destructive and can generate phenomics data at predetermined intervals (Klukas et al 2014; Chen et al 2014; Junker et al 2015). This is an indispensable tool in studying complex traits, thus supporting the discovery of dynamically controlled genetic factors in GWAs.

CONCLUSION

This study points out two additional possible answers to the missing heritability; i.e. CNV/PAV and dynamic controlled genetic factors. Though genome-wide association mapping approach that can account for these two genetic variants is anticipated to be highly successful in bringing genotype-phenotype existing gap, it faces some constraints. Insights into how genetic information in CNV/PAV will be translated into the genetic variability of complex traits are lacking. The genomic architectural complexity implies that in order to be able to detect CNV effects through association testing in larger populations, CNV endpoints need to be precisely delineated to assess information potentially masked by complex CNV architecture, such smaller CNV nested within larger CNV or overlapping CNV regions. Overlapping CNV regions result from inter-lineage variation, which was found to be very frequent in this study. A promising approach might therefore be to investigate the genetic basis of intermediate phenotypes with lower genetic complexity, such as yield components or metabolites, and link these results back to the complex trait of interest. Population and theoretical genetics approaches may hold the key to finding the missing heritability. However, incorporation of CNV/PAV and dynamically acting genetic factors in GWAS is anticipated to results in better estimation of heritability of complex traits. On the other hand, though automated high-throughput phenotyping platforms are indispensable tool in studying complex traits, the platforms are expensive to establish and thus limited to only a few experiments. Therefore, bridging this apparent genotype–phenotype gap remains a big challenge.

ACKNOWLEDGMENTS

The author acknowledges the technical support of Ingo Muecke and Beatrice Knüpffer, (Leibniz Institute of Plant Genetics and Crop Plant Research) for their help with the glasshouse and laboratory experiments. Furthermore, I thank Thomas Altmann (Leibniz Institute of Plant Genetics and Crop Plant Research) for his support throughout this project. This work was supported by grants from the Federal Ministry of Education and Research of Germany (BMBF, 0315461C).

SUPPORTING INFORMATION

Table S1: Candidate genes residing in the dynamic QTL regions, include the gene harbouring the SNP or within 55 kb up- or downstream the SNP marker. The panel analysed had a linkage decay of 55 kb.

REFERENCES

Alkan, C., Kidd, J.M., Marques-Bonet, T., Aksay, G., Antonacci, F., Hormozdiari, F., et al. 2009. Personalized copy number and segmental duplication maps using next-generation sequencing. Nature Genet 41:1061–1067

Bac-Molenaar, J.A., Vreugdenhil, D., Granier, C. and Keurentjes, J.J.B. 2015. Genome-wide association mapping of growth dynamics detects time-specific and general quantitative trait loci. Journal of Experimental Botany: doi:10.1093/jxb/erv176

Bellucci, A., Torp, A.M., Bruun, S., Magid, J., Andersen, S.B. and Rasmussen, S.K. 2015. Association Mapping in Scandinavian Winter Wheat for Yield, Plant Height, and Traits Important for Second-Generation Bioethanol Production. Front. Plant Sci. 6: 1046.

Belo', A., Beatty, M.K., Hondred, D., Fengler., K.A., Li, B. and Rafalski, A. 2010. Allelic genome structural variations in maize detected by array comparative genome hybridisation. Theor Appl Genet 120:355-367

Bergelson, J. and Roux, F. 2010. Identifying the genetic basis of complex traits in Arabidopsis thaliana. Nat Rev Genet 11:867-879.

Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Sci 48:1649–1664.

Bian, J.M., He, H.H., Li, C.J., Shi, H., Zhu, C.L., Peng, X.S., Fu, J.R., et al. 2013. Identification and validation of a new grain weight QTL in rice. Genet. Mol. Res. 12 (4): 5623-5633

Buckler, E.S., Holland, J.B., Bradbury, P.J., Acharya, C.B., Brown, P.J., et al. 2009. The genetic architecture of maize flowering time. Science 325:714-718

Busemeyer, L., Ruckelshausen, A., Möller, K., Melchinger, A.E., Alheit, K.V., Maurer, H.P., Hahn, V., Weissmann, E.A., Reif, J.C., Wuerschum, T. 2013. Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. Sci Rep 3:2442

Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, A. and Klukas, C. 2014. Dissecting the Phenotypic Components of Crop Plant Growthand Drought Responses Based on High-Throughput Image Analysis. Plant Cell. doi:10.1105/tpc.114.129601.

Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C, Campbell, P., Fitzgerald, T., Hu, M., Ihm, C.H., Kristiansson, K., Macarthur, D.G., Macdonald, J.R., Onyiah, I., Pang, A.W., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J; Wellcome Trust Case Control Consortium, Tyler-Smith, C., Carter, N.P., Lee, C., Scherer, S.W. and Hurles, M.E. 2010. Origins and functional impact of copy number variation in the human genome. Nature. 2010; 464:704-12

Diez, E., et al. 2003. Birc1e is the gene within the Lgn1 locus associated with resistance to Legionella pneumophila. Nature Genet 33:55–60.

Diskin, S.J., Hou, C., Glessner, J.T., Attiyeh, E.F., Laudenslager, M., Bosse, K., Cole, K., et al. 2009. Copy number variation at 1q21.1 associated with neuroblastoma. Nature 459, 987-991

Eihler, E.E., Flint, J., Gibson, G., Kong, A., Leal, S.M., Moore, J.H. and Nadeau, J.H. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet 11:446-450.

Fubank, R.T. and Tester, M. 2011. Phenomics—technologies to relieve the phenotyping bottleneck. Trends Plant Sci. 16: 635–644.

Gnal, M.W., Durstewitz, G., Polley, A., Bernard, A.L., Buckler E.S., Charcosset A., Clarke J.D., et al. 2011. A Large Maize (Zea mays L.) SNP Genotyping Array: Development and Germplasm Genotyping, and Genetic Mapping to Compare with the B73 Reference Genome. Plos ONE 6 (12) e28334

Ha, Z.F., Li, X.H., Liu, X.L., Xie, C.X., Li, M., Zhang, D. and Zhang, S. 2011. Meta-analysis of constitutive and adaptive QTL for drought tolerance in maize. Euphytica 174: 165-177

Hnrichsen, C.N., Chaignat, E., Reymond, A. 2009. Copy number variants, diseases and gene expression. Hum Mol Genet 18: R1–R8.

Holm, S. 1979. A simple sequentially rejective multiple test procedure. Scandinavian journal of statistics, 65-70.

Jiang, Y., Zhao, Y., Rodemann, B., Plieske, J., Kollers, S., Korzun, V., Ebmeyer, E., Argillier, O., Hinze, M., Ling, J., Röder, M.S., Ganal, M.W., Mette, M.F. and Reif, J.C. 2014. Potential and limits to

unravel the genetic architecture and predict the variation of Fusarium head blight resistance in European winter wheat (Triticum aestivum L.). Heredity doi:10.1038/hdy.2014.104

Johannes, F., Porcher, E., Teixeira, F.K., Saliba-Colombani, V., Simon, M., Agier, N., Bulski, A., Albuisson, J., Heredia, F., Audigier, P., Bouchez, D., Dillmann, C., Guerche, P., Hospital, F. and Colot, V. 2009. Assessing the impact of transgenerational epigenetic variation on complex traits. PLoS Genetics 2009, 5:e10000530.

Junker, A., Muraya, M.M., Weigelt-Fischer, K., Arana-Ceballos, F., Klukas, C., Melchinger, A.E., Meyer, R.C., Riewe, D. and Altmann, T. 2015. Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. Frontiers in Plant Sciences. 5(770):1-21. Doi: 10.3389/fpls.2014.00770

Klukas, C., Chen, D. and Pape, J.M. 2014. IAP: an open-source information system for high-throughput plant phenotyping. Plant Physiology: OI:10.1104/pp.113.233932

Liu, W., Gowda, M., Reif, J.C., Hahn, V., Ruckelshausen, A., Weissmann, E.A., Maurer, H.P. and Würschum, T. 2014. genetic dynamics underlying phenotypic development of biomass yield in triticale. BMC Genomics 15: 458

Luo, L., Boerwinkle, E. and Xiong, M. 2011. Association studies for next-generation sequencing. Genome Res 21:1099-1108

Mace, E.S., Buhariwalla, H.K. and Crouch, J.H. 2003. A high throughput DNA extraction protocol for molecular breeding programs. Plant Mol Biol Report 21:459a–459h

Manolio, T.A., Collins, F.S., Cox, N.J, Goldstein, D.B., Hindorff, L.A., Hunter, D.J., et al. 2009. Finding the missing heritability of complex diseases. Nature 461:747-753

Min-Oo, G, et al. 2003. Pyruvate kinase deficiency in mice protects against malaria. Nature Genet 35:357–362.

Paerson, A.H., Bowers, J.E., Bruggmann, R,, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, et al. 2009. The Sorghum bicolor genome and the diversification of grasses. Nature 457:551-556.

Pery, G.H., Yang, F., Marques-Bonet, T., Murphy, C., Fitzgerald T, Lee AS, Hyland C, Stone AC, Hurles ME, Tyler-Smith C, Eichler EE, Carter, N.P., Lee, C., Redon, R. 2008. Copy number variation and evolution in humans and chimpanzees. Genome Res. 18: 1618-1710

Plat, A., Vilhjálmsson, B.J. and Nordborg, M. 2010. Conditions under which genome-wide association studies will be positively misleading. Genetics 186:1045-1052.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei F, et al. 2009. The B73 maize genome: complexity, diversity and dynamics. Science 326: 1112-1115.

Seifert, M., Gohr, A., Strickert, M. and Grosse, I. 2012. Parsimonious Higher-Order Hidden Markov Models for Improved Array-CGH Analysis with Applications to Arabidopsis thaliana, PLoS Comp Biol 8: e1002286

Springer, N.M., Ying, K., Fu, Y., Ji, T., Yeh, C-T., Jia, Y., et al. 2009. Maize Inbreds Exhibit High Levels of Copy Number Variation (CNV) and Presence/Absence Variation (PAV) in Genome Content Plos Genetics 5 (11)

Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O.P., Ingason, A. and Steinberg, S. et al. 2008 Large recurrent microdeletions associated with schizophrenia. Nature 455, 232–237.

Stich, B, Melchinger A. 2010. An introduction to association mapping in plants. CAB Reviews 5:1–9.

Storey, JD, Akey JM, Kruglyak L 2005 Multiple locus linkage analysis of genome-wide expression in yeast. PLoS Biol, 3:e267.

Swanson-Wagner, RA, Eichten SR, Kumari S, Stein JC, Ware D, Springer N.M. 2010. Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res. 20:1689-1699

The International Schizophrenia Consortium. 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 455, 237–241.

Valdar, W., et al. 2006. Genome-wide genetic association of complex traits in heterogeneous stock mice. Nature Genet 38:879–887.

Wood, A.R., Hernandez, D.G, Nalls, M.A., Yaghootkar, H., et al. 2011. Allelic heterogeneity and more detailed analyses of known loci explain additional phenotypic variation and reveal complex patterns of association. Hum Mol Genet, 20:4082-4092.

Wurschum, T, Liu, W., Busemeyer, L., Tucker, M., Reif, J., Weissmann, E., Hahn, V., Ruckelshausen, A. and Maurer, H. 2014. Mapping dynamic QTL for plant height in triticale. BMC Genetics 15, 59

Würschum, T., Maurer, H.P., Schulz, B., Möhring, J. and Reif. J.C. 2011. Genome-wide association mapping reveals epistasis and genetic interaction networks in sugar beet. Theoretical and Applied Genetics 123:109-118.

Yu, J. and Buckler, E.S. 2006. Genetic association mapping and genome organization of maize. Current Opinion in Biotechnology 17:155–160.

Yu, P., Wang, C., Xu, Q., Feng, Y., Yuan, X., Yu, H., Wang, Y., Tang, S. and Wei, X. 2011 Detection of copy number variations in rice using array-based comparative genomic hybridization BMC Genomics12:372

Zhang, F., Gu, W., Hurles, M.E. and Lupski, J.R. 2009. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 10: 451–481

Zhang, J. 2003. Evolution by gene duplication: an update. Trends Ecol Evol. 18:292-298

Zhang, X., Cal, A.J. and Borevitz, J.O .2011. Genetic architecture of regulatory variation in Arabidopsis thaliana. Genome Res, 21:725-733.

Zheng, Z.P., Liu, X.H. 2013. Genetic analysis of agronomic traits associated with plant architecture by QTL mapping in maize. Genet. Mol. Res. 12 2: 1243-1253

Zhu, J., Sanborn, J.Z., Diekhans, M., Craig, B,. Lowe, C.B., Pringle, T.H. and Haussler, D. 2007. Comparative genomics search for losses of long-established genes on the human lineage. PLoS Comput Biol. 3:e247

Zu, J. and Li, J. 2014. Molecular dissection of complex agronomic traits of rice: a team effort by Chinese scientists in recent years. Natl Sci Rev 1