Analysis of small RNA expression in the context of heterosis formation in *Zea mays* L.

Dissertation

with the aim of achieving a doctoral degree at the Faculty of Mathematics, Informatics and Natural Sciences Department of Biology of the University of Hamburg

submitted by

Felix Seifert

2014 in Hamburg

Supervisor: Dr. Stefan Scholten 1st Examiner: Dr. Stefan Scholten 2nd Examiner: Prof. Dr. Arp Schnittger

Day of disputation: 31.10.2014

This work has been generated from May 2011 until August 2014 in the lab of Dr. Stefan Scholten in the group Developmental Biology and Biotechnology in the department of Biology at the University of Hamburg, Biocenter Klein Flottbek.

Oral and poster presentations, covering parts of this work:

Plant Biology 2013, 20.07. - 24.07.2013, Providence, RI, USA, Poster presentation

Post-transcriptional gene-regulation in plants, 25.07. -26.07.2013, Providence, RI, USA, Poster presentation

56th Maize Genetics Conference, 13.03. - 16.03.2014, Beijing, China, Oral presentation

Publications prepared during graduation and at least partly covered in this work:

Seifert, F., Thiemann, A., Pospisil, H., Scholten, S. (2012) Re-annotation of the maize oligonucleotide array. *Maydica* 57, 49-55

Scholten, S., Thiemann, A., Seifert, F., Frisch, M., Melchinger, A. E. (2013) Vorhersage von Hybridmerkmalen, filed patent DE 10 2013 111 980.8

Thiemann, A., Fu, J., Seifert, F., Grant-Downton, R. T., Schrag, T. A., Pospisil, H., Frisch, M., Melchinger, A. E., Scholten, S. (2014) Genome-wide meta-analysis of maize heterosis reveals the potential role of additive gene expression at pericentromeric loci. *BMC Plant Biology* 14:88

Seifert, F., Thiemann, A., Grant-Downton, R., Edelmann, S., Schrag, T., Gutierrez-Marcos, J.F., Frisch, M., Dickinson, H.G., Melchinger, A.E., Scholten, S. Pericentromeric 22-nt small RNAs constitute negative components in heterosis formation of maize. (submitted manuscript)

I. Table of Contents

I.	Table of Contents	IV
II.	Index of Figures	VII
III.	Index of Tables	X
IV.	Index of Formulas	XII
V.	Abbreviation index	XIII
VI.	Unit index	xv
1.	Introduction	1
1.1	Heterosis	1
1.2	Small RNAs	5
1.3	Viral gene silencing suppressors	8
1.4	Objectives	11
n	Material and Mathada	10
Z.		12
2.1	Materials	12
	2.1.1 Software	12
	2.1.2 Plant material, growth and phenotyping	13
	2.1.2.2 sRNA transcriptome analysis in inbred lines from a breeding population and hybrids of different heterosis-levels	
2.2	Molecularbiological methods	16
	2.2.1 RNA isolation	
	2.2.2 Small RNA sequencing	
2.3	Computational methods	
	2.3.1 Re-annotation of the 46k-maizearray	17
	2.3.1.1 Oligonucleotides localization and target gene identification	17
	2.3.1.2 Functional annotation of target genes	18
	2.3.2 Sequencing data processing	19
	2.3.2.1 sRNA sequencing data processing	
	2.3.2.2 sRNA sequencing data normalization	19
	2.3.2.9 P1/HC-Pro gang silencing suppressor experiment	ںے۔۔۔۔۔۔ مر
	2.3.3.1 Calculation of best parent heterosis increase	20
	2.3.3.2 sRNA population complexity analysis	20
	2.3.3.3 sRNA length distribution analysis	21
	2.3.3.4 Analysis for differentially expressed miRNAs	21

	2.3.4 SRNA transcriptome analysis in inbred lines from a breeding population	on Od
	and hybrids of different neterosis-levels	.21
	2.3.4.1 Sequence mapping of sRNA sequences	.21
	2.3.4.2 Annotation of sRNA sequences	.22
	2.3.4.3 sRNA differential expression analysis	.23
	2.3.4.4 sRNA population differential expression distance calculation	.24
	2.3.4.5 sRNA population diversity analysis	.24
	2.3.4.6 Inbred line differentially expressed sRNA population diversity	
	comparison and grouping	.25
	2.3.4.7 Association of inbred parent sRNA expression with hybrid trait-valu	ies25
	2.3.4.8 Correlation analysis of sRNA-based distance of inbred parents and	1
	heterosis	.27
	2.3.4.9 sRNAome-based prediction of hybrid trait values	.27
	2.3.4.10 ha-sRNA length distribution enrichment analysis	.28
	2.3.4.11 ha-sRNA annotation distribution analysis	.28
	2.3.4.12 Distribution of ha-sRNAs at and around gene loci	.28
	2.3.4.13 Enrichment analysis for ha-sRNAs for repeat super-families and	
	families	.28
	2.3.4.14 Genome-wide enrichment analysis for ha-sRNAs	.29
	2.3.4.15 ha-sRNA expression pattern analysis	.29
	2.3.4.16 ha-sRNA target prediction	.29
	2.3.5 Analysis of effects on the transcriptome mediated by ha-sRNAs	.30
	2.3.5.1 Correlation of ha-sRNA expression and microarray transcriptome	
	expression data	.30
	2.3.5.2 sRNA differential expression analysis	.30
	2.2.5.2. DNA and differential expression enclusion	~ ~
	2.3.5.3 RINASEQ differential expression analysis	.30
	2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed	.30
	2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions	.30 .31
2	2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions	.30 .31
3.	2.3.5.3 RNAseq differential expression analysis 2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions	.30 .31 .32
3. 3.1	2.3.5.3 RNAseq differential expression analysis 2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions Results Molecularbiological analyses	.30 .31 .32 .32
3. 3.1	 2.3.5.3 RNAseq differential expression analysis	.30 .31 .32 .32 .32
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis	.30 .31 .32 .32 .32 .32
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis	.30 .31 .32 .32 .32 .32
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis	.30 .31 .32 .32 .32 .32 .32
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis	.30 .31 .32 .32 .32 .32 .32 .32 .33
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis	.30 .31 .32 .32 .32 .32 .32 .32 .33 .33
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis	.30 .31 .32 .32 .32 .32 .32 .32 .33 .33 .34
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis 2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions Results Molecularbiological analyses 3.1.1 Small RNA sequencing Computational analyses 3.2.1 Re-annotation of the 46k-maizearray 3.2.1.1 Oligonucleotide localization and target gene identification 3.2.1.2 Functional annotation of target genes 3.2.2 Sequencing data processing 3.2.2 sRNA sequencing data normalization 	.30 .31 .32 .32 .32 .32 .32 .33 .33 .34 .35
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis 2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions Results Molecularbiological analyses 3.1.1 Small RNA sequencing Computational analyses 3.2.1 Re-annotation of the 46k-maizearray 3.2.1.1 Oligonucleotide localization and target gene identification 3.2.1.2 Functional annotation of target genes 3.2.2 Sequencing data processing 3.2.2.1 sRNA sequencing data processing 3.2.2.2 sRNA sequencing data normalization 	.30 .31 .32 .32 .32 .32 .32 .32 .33 .33 .34 .35 .26
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis 2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions Results Molecularbiological analyses 3.1.1 Small RNA sequencing Computational analyses 3.2.1 Re-annotation of the 46k-maizearray 3.2.1.1 Oligonucleotide localization and target gene identification 3.2.1.2 Functional annotation of target genes 3.2.2 Sequencing data processing 3.2.2.1 sRNA sequencing data processing 3.2.2.2 sRNA sequencing data normalization 3.2.3 P1/HC-Pro gene silencing suppressor experiment 3.2.3 1 Calculation of best parent betarasis increase 	.30 .31 .32 .32 .32 .32 .32 .33 .33 .33 .34 .35 .36 .36
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis 2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions Results Molecularbiological analyses 3.1.1 Small RNA sequencing Computational analyses 3.2.1 Re-annotation of the 46k-maizearray 3.2.1.1 Oligonucleotide localization and target gene identification 3.2.1.2 Functional annotation of target genes 3.2.2 Sequencing data processing 3.2.2.1 sRNA sequencing data processing 3.2.2.2 sRNA sequencing data normalization 3.2.3 P1/HC-Pro gene silencing suppressor experiment 3.2.3.1 Calculation of best-parent heterosis increase 3.2.3 a RNA population complexity analysis 	.30 .31 .32 .32 .32 .32 .32 .33 .33 .34 .35 .36 .36 .37
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis 2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions Results Molecularbiological analyses 3.1.1 Small RNA sequencing Computational analyses 3.2.1 Re-annotation of the 46k-maizearray 3.2.1.1 Oligonucleotide localization and target gene identification 3.2.1.2 Functional annotation of target genes 3.2.2 Sequencing data processing 3.2.2 sRNA sequencing data normalization 3.2.3 P1/HC-Pro gene silencing suppressor experiment 3.2.3 sRNA population complexity analysis 	.30 .31 .32 .32 .32 .32 .32 .32 .32 .33 .33 .34 .35 .36 .36 .37 .37
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis 2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions Results Molecularbiological analyses 3.1.1 Small RNA sequencing. Computational analyses 3.2.1 Re-annotation of the 46k-maizearray 3.2.1.1 Oligonucleotide localization and target gene identification 3.2.1.2 Functional annotation of target genes 3.2.2 Sequencing data processing 3.2.2.1 sRNA sequencing data processing 3.2.2.2 sRNA sequencing data normalization 3.2.3 P1/HC-Pro gene silencing suppressor experiment 3.2.3.1 Calculation of best-parent heterosis increase 3.2.3 sRNA length distribution analysis 3.2.3 sRNA length distribution analysis 	.30 .31 .32 .32 .32 .32 .32 .33 .33 .33 .34 .35 .36 .36 .37 .37 .38
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis 2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions Results Molecularbiological analyses 3.1.1 Small RNA sequencing Computational analyses 3.2.1 Re-annotation of the 46k-maizearray 3.2.1.1 Oligonucleotide localization and target gene identification 3.2.1.2 Functional annotation of target genes 3.2.2 Sequencing data processing 3.2.2.1 sRNA sequencing data processing 3.2.2.2 sRNA sequencing data normalization 3.2.3 P1/HC-Pro gene silencing suppressor experiment 3.2.3.3 sRNA length distribution analysis 3.2.3.4 Analysis for differentially expressed miRNAs 	.30 .31 .32 .32 .32 .32 .32 .33 .33 .34 .35 .36 .37 .37 .38
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis 2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions Results Molecularbiological analyses 3.1.1 Small RNA sequencing Computational analyses 3.2.1 Re-annotation of the 46k-maizearray 3.2.1.1 Oligonucleotide localization and target gene identification 3.2.1.2 Functional annotation of target genes 3.2.2 Sequencing data processing 3.2.2.1 sRNA sequencing data processing 3.2.2.2 sRNA sequencing data normalization 3.2.3 P1/HC-Pro gene silencing suppressor experiment 3.2.3.3 sRNA length distribution analysis 3.2.3.4 Analysis for differentially expressed miRNAs 3.2.4 sRNA/trait-association experiment 	.30 .31 .32 .32 .32 .32 .32 .32 .33 .34 .35 .36 .37 .38 .39
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis 2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions. Results Molecularbiological analyses 3.1.1 Small RNA sequencing. Computational analyses 3.2.1 Re-annotation of the 46k-maizearray 3.2.1.1 Oligonucleotide localization and target gene identification 3.2.1.2 Functional annotation of target genes 3.2.2 Sequencing data processing 3.2.2.1 sRNA sequencing data processing 3.2.2 sRNA sequencing data normalization 3.2.3 P1/HC-Pro gene silencing suppressor experiment 3.2.3 sRNA population complexity analysis 3.2.3 sRNA length distribution analysis 3.2.3 sRNA length distribution analysis 3.2.4 sRNA/trait-association experiment 3.2.4 sRNA/trait-association experiment 	.30 .31 .32 .32 .32 .32 .32 .33 .33 .34 .35 .36 .37 .38 .39 .39
3. 3.1 3.2	 2.3.5.3 RNAseq differential expression analysis 2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions Results Molecularbiological analyses 3.1.1 Small RNA sequencing. Computational analyses 3.2.1 Re-annotation of the 46k-maizearray 3.2.1.1 Oligonucleotide localization and target gene identification	.30 .31 .32 .32 .32 .32 .32 .33 .33 .33 .34 .35 .36 .37 .38 .39 .39 .40

	3.2.4 3.2.4	 4.4 Association of inbred parent sRNA expression with hybrid trait-value 5.5 Correlation analysis of sRNA-based distance of inbred parents a batarosis 	alues42 nd
	324	helerosis 6 sRNAome-based prediction of hybrid trait values	43 45
	3.2.	7 ha-sRNA length distribution enrichment analysis	46
	3.2.4	8. Genome-wide enrichment analysis for ha-sRNAs	48
	3.2.4	.9 ha-sRNA annotation distribution analysis	49
	3.2.4	.10 Distribution of ha-sRNAs at and around gene loci	50
	3.2.4	1.11 Enrichment/depletion analysis for ha-sRNAs for repeat super-	
		families and families	50
	3.2.4	12 ha-sRNA expression pattern analysis	56
	3.2.5	Analysis of effects on the transcriptome mediated by ha-sRNAs	57
	3.2.5	5.1 ha-sRNA target prediction	57
	3.2.5	5.2 Correlation of ha-sRNA with microarray expression data	57
	3.2.5	5.3 sRNA differential expression analysis	58
	3.2.5	5.4 RNAseq differential expression analysis	59
	3.2.5	5.5 Enrichment analysis for ha-sRNAs at differentially expressed	
		transcript genome regions	59
4.	Disc	ussion	60
4.1	Re-a	nnotation of the 46k-maizearray	60
4.2	sRN	A sequencing data normalization, the neglected factor	61
4.3	Fund	tional identification of aDNA offects on betaracia formation usi	ina a
	vira	al gene silencing suppressor	63
4.4	vira sRN and	A transcriptome analysis in inbred lines from a breeding popula hybrids of different heterosis-levels	ation 65
4.4	vira sRN and 4.4.1	A transcriptome analysis in inbred lines from a breeding popula hybrids of different heterosis-levels sRNA transcriptomes reflect the genetic diversity of maize inbred lines	63 ation 65 nes65
4.4	vira sRN and 4.4.1 4.4.2	A transcriptome analysis in inbred lines from a breeding popula hybrids of different heterosis-levels sRNA transcriptomes reflect the genetic diversity of maize inbred lin ha-sRNAs act antagonistically in heterosis formation	ation 65 65 68
4.4	vira sRN and 4.4.1 4.4.2 4.4.3	A transcriptome analysis in inbred lines from a breeding popula A hybrids of different heterosis-levels sRNA transcriptomes reflect the genetic diversity of maize inbred lin ha-sRNAs act antagonistically in heterosis formation ha-sRNA characterization reveals enrichment of specific sRNA feat	ation 65 65 68 ures69
4.4	vira sRN and 4.4.1 4.4.2 4.4.3 4.4.3	A transcriptome analysis in inbred lines from a breeding popula hybrids of different heterosis-levels sRNA transcriptomes reflect the genetic diversity of maize inbred lin ha-sRNAs act antagonistically in heterosis formation ha-sRNA characterization reveals enrichment of specific sRNA feat	ation 65 nes65 68 cures69
4.4	vira sRN and 4.4.1 4.4.2 4.4.3 4.4.4	A transcriptome analysis in inbred lines from a breeding popula hybrids of different heterosis-levels sRNA transcriptomes reflect the genetic diversity of maize inbred lin ha-sRNAs act antagonistically in heterosis formation ha-sRNA characterization reveals enrichment of specific sRNA feat Analysis of effects on the transcriptome mediated by ha-sRNAs	ation 65 nes65 68 ures69 76
4.4	vira sRN and 4.4.1 4.4.2 4.4.3 4.4.3 4.4.4 4.4.5	A transcriptome analysis in inbred lines from a breeding popula hybrids of different heterosis-levels sRNA transcriptomes reflect the genetic diversity of maize inbred lin ha-sRNAs act antagonistically in heterosis formation ha-sRNA characterization reveals enrichment of specific sRNA feat Analysis of effects on the transcriptome mediated by ha-sRNAs ha-sRNAs hold the potential for the prediction of hybrid traits and implications for future plant breeding.	ation 65 nes65 68 aures69 76
4.4	vira sRN and 4.4.1 4.4.2 4.4.3 4.4.3 4.4.4 4.4.5	A transcriptome analysis in inbred lines from a breeding popula hybrids of different heterosis-levels sRNA transcriptomes reflect the genetic diversity of maize inbred lin ha-sRNAs act antagonistically in heterosis formation ha-sRNA characterization reveals enrichment of specific sRNA feat Analysis of effects on the transcriptome mediated by ha-sRNAs ha-sRNAs hold the potential for the prediction of hybrid traits and implications for future plant breeding	ation 65 nes65 68 sures69 76 77
4.4 5.	vira sRN and 4.4.1 4.4.2 4.4.3 4.4.4 4.4.5 Pers	A transcriptome analysis in inbred lines from a breeding popula hybrids of different heterosis-levels sRNA transcriptomes reflect the genetic diversity of maize inbred lin ha-sRNAs act antagonistically in heterosis formation ha-sRNA characterization reveals enrichment of specific sRNA feat Analysis of effects on the transcriptome mediated by ha-sRNAs ha-sRNAs hold the potential for the prediction of hybrid traits and implications for future plant breeding	ation 65 nes65 68 sures69 76 77 77
4.4 5. VII.	vira sRN and 4.4.1 4.4.2 4.4.3 4.4.4 4.4.5 Pers Abst	A transcriptome analysis in inbred lines from a breeding population of sRNA effects on neterosis formation using a gene silencing suppressor	63 ation 65 nes65 68 aures69 76 77 80 XVIII
4.4 5. VII. VIII.	vira sRN and 4.4.1 4.4.2 4.4.3 4.4.4 4.4.5 Pers Abst Zusa	A transcriptome analysis in inbred lines from a breeding popular A transcriptome analysis in inbred lines from a breeding popular hybrids of different heterosis-levels sRNA transcriptomes reflect the genetic diversity of maize inbred line ha-sRNAs act antagonistically in heterosis formation	63 ation 65 nes65 68 cures69 76 77 77 80 XVIII XIX
4.4 5. VII. VIII. IX.	vira sRN and 4.4.1 4.4.2 4.4.3 4.4.4 4.4.5 Pers Abst Zusa Refe	A transcriptome analysis in inbred lines from a breeding popula hybrids of different heterosis-levels sRNA transcriptomes reflect the genetic diversity of maize inbred line ha-sRNAs act antagonistically in heterosis formation	63 ation 65 nes65 68 aures69 76 77 77 80 XVIII XIX XX
4.4 5. VII. VIII. IX. X.	vira sRN and 4.4.1 4.4.2 4.4.3 4.4.4 4.4.5 Pers Abst Zusa Refe Appe	A transcriptome analysis in inbred lines from a breeding popula I hybrids of different heterosis-levels sRNA transcriptomes reflect the genetic diversity of maize inbred lin ha-sRNAs act antagonistically in heterosis formation	63 ation 65 nes65 68 aures69 76 77 80 XVIII XIX XX XXVI
4.4 5. VII. VIII. IX. X. XI.	vira sRN and 4.4.1 4.4.2 4.4.3 4.4.4 4.4.5 Pers Abst Zusa Refe Appe Decl	A transcriptome analysis in inbred lines from a breeding popula hybrids of different heterosis-levels sRNA transcriptomes reflect the genetic diversity of maize inbred line ha-sRNAs act antagonistically in heterosis formation	63 ation 65 nes65 68 aures69 76 77 80 XVIII XIX XX XXVI LVIII
4.4 5. VII. VIII. IX. X. XI. XII.	vira sRN and 4.4.1 4.4.2 4.4.3 4.4.4 4.4.5 Pers Abst Zusa Refe Appe Decl Conf	A transcriptome analysis in inbred lines from a breeding popula hybrids of different heterosis-levels sRNA transcriptomes reflect the genetic diversity of maize inbred line ha-sRNAs act antagonistically in heterosis formation	63 ation 65 nes65 68 aures69 76 77 80 XVIII XIX XIX XX XXVI LVIII LVIII LIX

II. Index of Figures

Figure 1: Heterosis in maize1
Figure 2: Endogenous sRNA biogenesis pathways in plants (Pumplin & Voinnet 2013)7
Figure 3: Model of the function of various viral gene silencing suppressors (yellow hexagons) on plants gene silencing components (Peláez & Sanchez 2013)
Figure 4: Schematic drawing of the P1/HC-Pro construct
Figure 5: maizearray re-annotation procedure
Figure 6: sRNA length distribution of raw total read counts of the reference inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of Hohenheim
Figure 7: sRNA length distribution of distinct read counts of the reference inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of Hohenheim
Figure 8: Total sRNA read counts before (raw) and after quantile normalization of the reference inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of Hohenheim
Figure 9: sRNA length distribution of normalized total read counts [rpmqn] of the reference inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of Hohenheim
Figure 10: BPH for growth for induced/non-induced transgenic P1/HC-Pro hybrid plants 36
Figure 10: BPH for growth for induced/non-induced transgenic P1/HC-Pro hybrid plants
Figure 10: BPH for growth for induced/non-induced transgenic P1/HC-Pro hybrid plants
Figure 10: BPH for growth for induced/non-induced transgenic P1/HC-Pro hybrid plants
Figure 10: BPH for growth for induced/non-induced transgenic P1/HC-Pro hybrid plants 36 Figure 11: sRNA population complexity reduction analysis results 37 Figure 12: sRNA length distribution comparison analysis between induced and non-induced P1/HC-Pro hybrids 37 Figure 13: sRNA mapping count distribution of the reference inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of Hohenheim 39 Figure 14: Average inbred parent sRNA population diversity analysis from 98 inbred line pairs of two heterotic groups (Dent and Flint, 14x7 inbred combinations)
Figure 10: BPH for growth for induced/non-induced transgenic P1/HC-Pro hybrid plants
Figure 10: BPH for growth for induced/non-induced transgenic P1/HC-Pro hybrid plants 36 Figure 11: sRNA population complexity reduction analysis results 37 Figure 12: sRNA length distribution comparison analysis between induced and non-induced P1/HC-Pro hybrids 37 Figure 13: sRNA mapping count distribution of the reference inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of Hohenheim 39 Figure 14: Average inbred parent sRNA population diversity analysis from 98 inbred line pairs of two heterotic groups (Dent and Flint, 14x7 inbred combinations) 40 Figure 15: sRNA population diversity of the low heterotic hybrid S028xL024 and its parental inbred lines 41 Figure 16: sRNA population diversity of the intermediate heterotic hybrid S028xF039 and its parental inbred lines 41
Figure 10: BPH for growth for induced/non-induced transgenic P1/HC-Pro hybrid plants 36 Figure 11: sRNA population complexity reduction analysis results 37 Figure 12: sRNA length distribution comparison analysis between induced and non-induced P1/HC-Pro hybrids 37 Figure 13: sRNA mapping count distribution of the reference inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of Hohenheim 39 Figure 14: Average inbred parent sRNA population diversity analysis from 98 inbred line pairs of two heterotic groups (Dent and Flint, 14x7 inbred combinations). 40 Figure 15: sRNA population diversity of the low heterotic hybrid S028xL024 and its parental inbred lines 41 Figure 16: sRNA population diversity of the intermediate heterotic hybrid S028xF039 and its parental inbred lines 41 Figure 17: sRNA population diversity of the high heterotic hybrid P033xF047 and its parental inbred lines 41

Figure 19: Correlation of binary distance of differentially expressed sRNAs with heterosis of all 98 hybrids
Figure 20: Correlation of euclidean distance of differentially expressed sRNAs with heterosis of all 98 hybrids
Figure 21: Correlation of binary distance of positive ha-sRNAs with heterosis of all 98 hybrids44
Figure 22: Correlation of binary distance of negative ha-sRNAs with heterosis of all 98 hybrids
Figure 23: Correlation of binary distance of all ha-sRNAs with heterosis of all 98 hybrids 44
Figure 24: Prediction accuracy for ha-sRNA predictions of heterosis (MPH for GY)45
Figure 25: Prediction accuracy for ha-sRNA based predictions of hybrid performance (HP for GY)45
Figure 26: Length distribution and enrichment analysis of positive and negative ha-sRNAs 46
Figure 27: pos. ha-sRNA length distribution separately for mapping count groups
Figure 28: neg. ha-sRNA length distribution separately for mapping count groups
Figure 29: Genome-wide enrichment analysis for ha-sRNAs
Figure 30: Distribution of ha-sRNAs to annotation types
Figure 31: Relative distribution of ha-sRNAs at and adjacent to gene annotations
Figure 32: sRNA length distribution of ha-sRNAs to repeat super-families
Figure 33: sRNA length distribution of bootstrap sample average to repeat super-families51
Figure 34: ha-sRNA enrichment and depletion probabilities51
Figure 35: sRNA length distribution of pos. ha-sRNAs to repeat super-families
Figure 36: sRNA length distribution of neg. ha-sRNAs to repeat super-families52
Figure 37: pos. ha-sRNA enrichment and depletion probabilities to repeat super-families52
Figure 38: neg. ha-sRNA enrichment and depletion probabilities to repeat super-families52
Figure 39: sRNA length distribution of ha-sRNAs to highly expressed repeat families
Figure 40: ha-sRNA enrichment and depletion probabilities to repeat families53
Figure 41: sRNA length distribution of pos. ha-sRNAs to highly expressed repeat families 54
Figure 42: pos. ha-sRNA enrichment and depletion probabilities to repeat families
Figure 43: sRNA length distribution of neg. ha-sRNAs to highly expressed repeat families 55
Figure 44: neg. ha-sRNA enrichment and depletion probabilities to repeat families
Figure 45: Dominance/additivity (d/a) expression pattern analysis of differentially expressed sRNAs and ha-sRNAs

Figure 46: Expression pattern ratio for fold-change and presence-absence variation of 22 nt
and 24 nt ha-sRNAs based on 14x7 inbred combinations56
Figure 47: Length distribution of ha-sRNAs with expression significantly correlated to putative target transcripts
Figure 48: sRNA length distribution of ha-sRNAs present in the inbred lines B73 and Mo1758

III. Index of Tables

Table 1: Software used in this study 12
Table 2: Total and relative re-annotation results for the maize oligonucleotide array platforms 33
Table 3: Result summary of ha-sRNA target prediction with psRNATarget
Table 4: Distribution of ha-sRNAs of specific length to regions of differentially expressed genes
Appendix Table 1: Published experiments performed on either the 57K or 46K maizearray
Appendix Table 2: Hybrid trait values for MPH for GY [Mg ha-1 adjusted to 155 g kg-1 grain moisture] of 98 hybrids resulting from 14x7 inbred combinationsXXXVII
Appendix Table 3: Growth-rates of induced/non-induced transgenic P1/HC-Pro inbred and hybrid plants [cm/2 d]XXXVIII
Appendix Table 4: Raw and processed read counts from P1/HC-Pro hybrid sequencing .XXXIX
Appendix Table 5: Raw and processed sequence counts from inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of HohenheimXXXIX
Appendix Table 6: Raw and quantile normalized total read counts of the inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of HohenheimXL
Appendix Table 7: Results from sRNA population complexity analysis for induced/non-induced transgenic P1/HC-Pro plants (bootstrap analysis with 1000 runs)XLI
Appendix Table 8: sRNA length distribution of normalized read counts for induced/non-induced P1/HC-Pro transgenic hybridsXLI
Appendix Table 9: sRNA read mapping distribution to the B73 reference genome of the inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of Hohenheim
Appendix Table 10: Probabilities of differences in read mapping position counts between Flint and Dent lines as well as inbreds and hybrids
Appendix Table 11: sRNA population overlaps between inbred parents [%]XLIII
Appendix Table 12: ha-sRNA length distribution and enrichment probabilities XLIV
Appendix Table 13: sRNA length distribution of pos. ha-sRNA for mapping position count groups on the B73 reference genome
Appendix Table 14: sRNA length distribution of neg. ha-sRNAs for mapping position count groups on the B73 reference genome
Appendix Table 15: sRNA length distribution of pos. ha-sRNAs to annotation types XLV

Appendix Table 16: sRNA length distribution of neg. ha-sRNAs to annotation types XLVI
Appendix Table 17: sRNA length distribution of ha-sRNAs to repeat super-families XLVI
Appendix Table 18: Average sRNA length distribution to repeat super-families calculated from 1000 bootstrap runs
Appendix Table 19: sRNA length enrichment/depletion probabilities for ha-sRNA to repeat super-families
Appendix Table 20: sRNA length distribution of pos. ha-sRNAs to repeat super-families XLIX
Appendix Table 21: sRNA length distribution of neg. ha-sRNAs to repeat super-families XLIX
Appendix Table 22: sRNA length enrichment/depletion probabilities for pos. ha-sRNA to repeat super-familiesL
Appendix Table 23: sRNA length enrichment/depletion probabilities for neg. ha-sRNA to repeat super-familiesLI
Appendix Table 24: sRNA length distribution of ha-sRNAs to repeat familiesLII
Appendix Table 25: sRNA length enrichment/depletion probabilities for ha-sRNAs to repeat families
Appendix Table 26: sRNA length distribution of pos. ha-sRNAs to repeat familiesLIV
Appendix Table 27: sRNA length enrichment/depletion probabilities pos. ha-sRNA to repeat familiesLIV
Appendix Table 28: sRNA length distribution of neg. ha-sRNAs to repeat familiesLV
Appendix Table 29: sRNA length enrichment/depletion probabilities neg. ha-sRNA to repeat familiesLVI
Appendix Table 30: Length distribution of ha-sRNAs with expression significantly correlated to putative target transcriptsLVII
Appendix Table 31: sRNA length distribution of ha-sRNAs, present in the inbred lines B73 and Mo17LVII

IV. Index of Formulas

Formula 1: Differential expression calculation	23
Formula 2: High parent expression value	23
Formula 3: Low parent expression value	23
Formula 4: Binary distance calculation for a pair of inbred lines	24
Formula 5: Euclidean distance calculation for a pair of inbred lines	24
Formula 6: Association probability calculation	26
Formula 7: Binomial probability mass function	26
Formula 8: Sum of differentially expressed sRNAs in both groups	26
Formula 9: Expression parameters for positive association	26
Formula 10: Expression parameters for negative association	26
Formula 11: Combined binary distance of pos. and neg. associated sRNAs	27

V. Abbreviation index

ARGONAUTE
binary sequence alignment/ map format
bialaphos resistance
best parent heterosis
circa
copy DNA
coding sequence
Cauliflower mosaic virus
copy number variation
DICER-LIKE
Dexamethasone
deoxyribonucleic acid
dominance to additivity
days after germination
days after pollination
double-stranded RNA
for example
expressed sequence tag
first filial generation
false discovery rate
galactose
general combining ability
gene ontology
glucocorticoid receptor
β-glucoronidase

GY	grain yield
ha-sRNA	heterosis-associated sRNA
HST	HASTY
HC-Pro	helper component protein- ase
hc-siRNA	heterochromatic siRNA
HEN1	HUA enhancer 1
HP	hybrid performance
indels	insertions or deletions
L.	Linné
ImiRNA	long-miRNA
LTR	long terminal repeat
mRNA	messenger RNA
miRNA	micro RNA
MIR	miRNA gene
MPH	mid-parent heterosis
MS	Murashige-Skoog
NAT	natural antisense transcript
nat-siRNA	natural antisense transcript siRNA
neg.	negative
nr	non-redundant protein da- tabase
ORF	open reading frame
р	p-value
P1	proteinase 1
ΡΑΤ	phosphinothricin- acetyltransferase

PAV	presence-absence variation	SRA	sequence read archive
PMV	Potato mosaic virus	sRNA	small RNA
PCA	principal component analy-	siRNA	short interfering RNA
Pol	sis DNA-dependent RNA-	SPCSV	Sweet potato chlorotic stunt virus
	polymerase	SSR	simple sequence repeat
pos.	positive	ssRNA	single stranded RNA
pre-miRNA	miRNA precursor	TAS	ta-siRNA gene
pri-miRNA	primary miRNA	ta-siRNA	trans-acting siRNA
PTGS	post-transcriptional gene	TCP	Turnip crinkle virus
PVX	Potato virus X	TE	transposable element
r	correlation coefficient	TIR	terminal inverted repeat
' RAPD	random amplified polymor-	TMV	Tomato mosaic virus
	phic DNA	TSS	transcription start site
ra-siRNAs	repeat-associated siRNAs	TYLCV	Tomato yellow leaf curl vi- rus
RdDM	RNA-dependent DNA- methylation	UBI	ubiquitin-ribosomal protein
RDR	RNA-dependent RNA polymerase	VIGS	virus-induced gene silenc- ing
RISC	RNA-induced silencing	WGS	working gene set
	complex	WSMV	wheat streak mosaic virus
RNA	ribonucleic acid		
RNAi	RNA interference	Gene names	s are written in capital letters.
RNase	ribonuclease		
SAM	sequence alignment/map format		
sec-siRNAs	secondary siRNAs		
SGS	Suppressor of gene silenc- ing		
SNP	single nucleotide polymor- phism		

VI. Unit index

bp	base-pair
ha	hectar (0,01 km ²)
L	liter (10 cm3)
mol	molar mass
nt	nucleotide
rpm	read counts per million reads
rpmqn	read counts per million quantile normalized reads

This work is using SI units, all other units are listed in the unit index. It should be noted that Mg stands for mega gram not Magnesium.

1. Introduction

1.1 Heterosis

Heterosis or hybrid vigor describes the phenomenon of the increased phenotypic performance of heterozygous hybrid offspring in comparison to their homozygous parental inbred lines. The effects of inbreeding and crossbreeding on the phenotypic performance of various plants were already observed by Darwin (Darwin 1876). Shull (1908) made the observation in maize that inbred lines show a general decrease in vigor and yield while their resulting hybrids recover and even outperform the parental phenotypes (see Figure 1).



Figure 1: Heterosis in maize

The picture shows plants of the inbred lines UH301 (301x301, left) and UH005 (005x005, right) and their reciprocal hybrids (301x005, 005x301, two plants in the middle). The hybrids exhibit heterosis for plant height exceeding their parents (Meyer *et al.* 2007)

He later on founded the term heterosis as an abbreviation for "stimulation of heterozygosity" (Shull 1914). This superior phenotype of hybrids was demonstrated for various relevant traits, e.g. biomass, yield, growth rate, fertility, environmental adaptation, and abiotic/biotic stress resistance (Shull 1909, Duvick 1999, Flint-Garcia *et al.* 2009). F1 offspring obtained from the crossing of two homozygous inbred lines is of both high agronomic and economic value due to its uniformity. Thus Shull proposed a breeding concept based on inbred lines that later became the standard in corn-breeding programs (Shull 1909, Crow 1998). Since the 1930s about 60 % of the maize yield increase could be attributed to genetic improvements due to the change from open-

pollination to hybrid corn breeding (Cardwell 1982, Duvick 2005). It was furthermore shown that gains in yield were primarily obtained due to genetic improvements related to tolerance to biotic and abiotic stresses (Duvick *et al.* 2004).

Breeders aim to generate hybrids with high hybrid performance for specific traits. Heterotic groups were established for a systematic exploitation of heterosis in hybrid breeding by separation of the inbred lines into at least two genetically divergent germplasm populations (Reif et al. 2005). A heterotic group was defined by Melchinger & Gumber (1998) "as a group of related or unrelated genotypes from the same or different populations, which display similar combining ability and heterotic response where crossed with genotypes from other genetically distinct germplasm groups". The selection of optimal inbred lines from these heterotic groups for hybrid crosses is a crucial step in the breeding process. Testing all possible line combinations is intensive in both time and cost, thus various prediction approaches for the selection of the best crossing partners were developed based on genetic markers, e.g. amplified fragment length polymorphisms (AFLPs), random amplified polymorphic DNAs (RAPDs), simple sequence repeats (SSR), single nucleotide polymorphisms (SNPs), and transcriptome, proteome, or metabolome profiles (Reif et al. 2003, Schrag et al. 2006, Xie et al. 2006, Frisch et al. 2010, Riedelsheimer et al. 2013). Mechanistic insights into the establishment of heterosis will allow improving hybrid breeding. Despite its successful application in crop breeding and extensive investigation for over a century, all available hypotheses are not yet able to fully explain the genetic or molecular basis of heterosis (Schnable & Springer 2013). The dominance hypothesis defines heterosis as the complementation of slightly deleterious recessive alleles present in one parent by dominant alleles from the opposed parent (Bruce 1910, Crow 1948). Assuming that this hypothesis exclusively explains heterosis, it should be possible to generate an inbred line that contains all favorable alleles from a hybrid (Charlesworth & Willis 2009). The overdominance hypothesis assigns the superior performance of heterozygous F1 offspring in comparison to its homozygous parents to favorable allelic interactions (East 1936, Crow 1948), thus heterozygosity per se should result in heterosis. Pseudo-overdominance is a mimic of overdominance caused by repulsion-phase linkage of favorable and detrimental alleles (Moll et al. 1965). In contrast, the epistasis hypothesis states heterosis to arise from non-allelic genic interactions at multiple loci in hybrids (Goodnight 1999). Most of these models were proposed before molecular concepts of genetics were established and investigated. Although they are not understood to be mutually exclusive, these hypotheses were and still are unable to explain the complex interactions that result in heterosis, which is more likely assumed to depend on multiple mechanisms, including epigenetics (Crow 1998, Birchler *et al.* 2003, Swanson-Wagner *et al.* 2006). The molecular concept of heterosis provides two major opposing models, namely additive and non-additive gene expression. Additive (mid-parent) expression corresponds to the hybrid having the average expression of the two parents and is caused by combinatorial interactions of alleles of both parents. Non-additive expression itself is split in two patterns, namely dominance and over-/under-dominance. Dominant expression refers to the case that the expression in the hybrid equals that of one of its parents (high/low-parent-like), while over-/under-dominance means that the expression in the hybrid is higher or lower than the expression in its parents (above high parent/below low parent) respectively (Schnable & Springer 2013).

All hypotheses clearly define heterosis as a result of the interaction of the parental genomes in the progeny. This merging and coherent transformation of the genome, epigenome and transcriptome results in changes and establishment of new regulatory cascades and thereby results in altered gene expression patterns (Michalak 2009). It was shown that corresponding sequence regions in maize inbred lines differ notably due to single nucleotide polymorphisms (SNP), small insertions/deletions, copy number variation (CNV), and presence-absence variation (PAV) (Tenaillon et al. 2001, Song & Messing 2003, Springer et al. 2009, Hansey et al. 2012). The strong heterotic effects observed in maize are assumed to be caused by the high intraspecific divergence, resulting in wide genetic distances between maize inbred lines (Moll et al. 1965, Reif et al. 2003). It was shown that the genetic basis of heterosis depends on the trait and is not subject to a general underlying mechanism (e.g. heterozygosity) or single genes (Flint-Garcia et al. 2009, Schnable & Springer 2013). Differential gene expression between the parental inbred lines as well as their hybrids is supposed to be involved in formation of heterosis (Romagnoli et al. 1990, Tsaftaris 1995, Guo et al. 2006). Although heterosis-associated gene expression did not reveal key genes, it was shown that expression patterns correlated with heterosis (Parvez 2006). More significant expression differences were found between parental inbred lines than between reciprocal hybrids (Stupar & Springer 2006). Differential gene expression between different lines is caused by *cis*- and *trans*-regulation at transcription levels, epigenetic modifications and posttranscriptional adjustment (Yao *et al.* 2005, Song *et al.* 2007).

It was shown that DNA methylation patterns exhibit significant variation between different maize genetic materials and that they are involved in the regulation of gene expression. Thus they have been suggested to be involved in the development of heterosis (Tsaftaris 1995). Drastic changes in DNA methylation levels in the comparison of hybrids and their inbred parents gave further support to this hypothesis (Xiong et al. 1999). The identification of small RNAs (sRNAs) as mediators of DNA methylation (Baulcombe 2004) and changes in sRNA expression and populations, shown between inbred lines and their hybrids, resulted in the assumption that sRNAs are involved in the molecular mechanisms related to heterosis formation (Swanson-Wagner et al. 2006, Springer & Stupar 2007, Groszmann et al. 2011, Barber et al. 2012, He et al. 2013). A large number of studies revealed differences in sRNA expression between two inbred lines and their hybrid offspring in Arabidopsis (Ha et al. 2009, Groszmann et al. 2011, Li et al. 2012, Shen et al. 2012), maize (Barber et al. 2012, He et al. 2013), rice (Chen et al. 2010, He et al. 2010, Chodavarapu et al. 2012), and wheat (Kenan-Eichler et al. 2011). Although all these studies uncovered differences in the sRNA transcriptomes of inbred lines and hybrids, these studies were unable to answer the question if these differences are involved in the formation or an effect of heterosis.

1.2 Small RNAs

Small RNAs (sRNAs) are a grouping of small non-protein-coding RNAs that are involved in transcriptional, post-transcriptional, and translational gene-regulation, and in the modulation of the epigenetic landscape by various mechanisms (Baulcombe 2004, Chen 2009, Finnegan & Matzke 2003, Vance & Vaucheret 2001, Xie & Qi 2008, Castel & Martienssen 2013). sRNAs have a length of 20 nt to 24 nt bearing a 5' phosphate and are 2' O-methylated on their 3' terminal end (Ghildiyal & Zamore 2009, Axtell 2013, Li et al. 2005). sRNAs have common features in biogenesis and action with slight differences. The biogenesis relies on at least partially double-stranded RNAs (dsRNAs), which are recognized and cleaved by a dsRNA-specific protein of the DICER-LIKE (DCL) family, an RNase III family ribonuclease (Bernstein et al. 2001). The cleaved sRNA is incorporated into an RNA-induced silencing complex (RISC) containing a member of the Argonaute (AGO) protein family that performs repressive action on an RNA or DNA with sufficient complementarity to the sRNA (Thieme et al. 2012, Axtell 2013). sRNA-guided RISCs are involved in diverse biological functions, e.g. regulation of gene expression, silencing of transposable elements, viral defense and heterochromatin formation (Brodersen & Voinnet 2006). In plants, two major classes of sRNAs are known, namely micro RNAs (miRNAs) and short interfering RNAs (siRNAs).

miRNAs are a group of ~21 nt sRNAs defined by their biogenesis from single-stranded RNAs (ssRNAs) that are able to fold into a characteristic stem-loop secondary structure (Lau *et al.* 2001). miRNAs were primarily discovered in the nematode *Caenorhabditis elegans* followed by identifications of various other miRNAs with identical biogenesis in other animals and plants (Lee *et al.* 1993, Wightman *et al.* 1993, Lau *et al.* 2001, Lee *et al.* 2001, Llave *et al.* 2002a, Reinhart *et al.* 2002).

The biogenesis of plant miRNAs is carried out in multiple processing steps and starts with the transcription of the primary miRNA (pri-miRNA), typically by RNA polymerase II (Pol II). The lengths of pri-miRNA transcripts in maize range from 250 nt to 2000 nt. In all plants, they exhibit Pol II-features as 5' 7-methylguanylate cap, 3' polyadenylation tail, and intron splicing (Zhang *et al.* 2009). The pri-miRNA is able to fold into a stem-loop secondary structure that is recognized and processed by a DCL-protein, typically DCL1, by cleaving the pri-miRNA and thus generating the precursor miRNA (pre-miRNA). The pre-miRNAs, with a length of 50 nt to more than 350 nt, exhibit the characteristic stem-loop structure with high complementarity in the stem region (Bonnet *et al.*

2006). From the stem region of the pre-miRNA, the miRNA-duplex with a length of ~21 nt and 2 nt 3'-overhangs is excised by a DCL-protein, typically DCL1 (Park *et al.* 2002, Kurihara & Watanabe 2004, Czech & Hannon 2011). The 2'-OH at the 3' ends of the miRNA-duplex are methylated by HEN1 (Hua Enhancer 1) to prevent immediate degradation (Park *et al.* 2002, Yu *et al.* 2005). The miRNA-duplex is exported from the nucleus into the cytoplasm by HST (HASTY) and the miRNA-strand of the miRNA-duplex is loaded into an AGO1 containing RISC, while the opposite strand is being degraded (Baumberger & Baulcombe 2005, Park *et al.* 2005, Chen 2009). The RISC is guided to an open reading frame (ORF) of a messenger RNA (mRNA) transcript that is perfectly or nearly perfectly complementary to the miRNA and performs post-transcriptional regulation of gene expression by either transcript degradation or translational inhibition (Carrington & Ambros 2003, Zhang *et al.* 2009). In plants in almost all cases perfect pairing of miRNA and target transcript leads to degradation at the phosphodiester bond opposite the 10th and 11th nucleotide of the miRNA (Bartel 2004, Llave *et al.* 2002a).

Although some miRNA families as well as their target transcripts were shown to be conserved over large evolutionary scales, the majority of miRNAs was demonstrated to be lineage-specific (Axtell & Bowman 2008, Cuperus *et al.* 2011). The majority of miRNAs is targeting transcription factors involved in development or abiotic/biotic stress response (Bonnet *et al.* 2006, Jones-Rhoades *et al.* 2006). Few miRNAs are involved in trans-acting siRNA (ta-siRNA) processing by setting the register of the first cleavage site. Alternative processing of pre-miRNAs by DCL3 instead of DCL1 results in 24 nt sRNAs that are incorporated into an AGO4 containing RISC guiding DNA-methylation of complementary regions. These 24 nt sRNAs are called long miRNAs (ImiRNAs) or miRNA gene (MIR)-derived siRNAs and are assumed to be transcribed by Pol IV rather than Pol II (Wu *et al.* 2010, Chellappan *et al.* 2010).

The effect of post-transcriptional gene silencing guided by siRNAs was primarily observed in an experiment aiming to produce deep purple petunia flowers by overexpression of a chalcone synthase coding gene, unexpectedly resulting in white flowers (Napoli *et al.* 1990, van der Krol *et al.* 1990). The molecular mechanism of RNA interference (RNAi) was first discovered in *Caenorhabditis elegans* (Fire *et al.* 1998). RNAi became thereafter a powerful tool for reverse-genetics experiments to study gene functions by reducing transcript expression levels (Dorsett & Tuschl 2004). siRNAs are assumed to have been evolved as a host defense pathway against foreign or transposon derived nucleic acids (Finnegan & Matzke 2003). siRNAs are classified into various subgroups mainly by their biogenesis. These classes are: heterochromatic or repeat-associated siRNAs, natural antisense transcript siRNAs, secondary siRNAs and trans-acting siRNAs (Axtell 2013) (see Figure 2).



Figure 2: Endogenous sRNA biogenesis pathways in plants (Pumplin & Voinnet 2013) a) miRNA, b) repeat-associated siRNAs, c) trans-acting-siRNAs

Heterochromatic siRNAs (hc-siRNAs), also called repeat-associated siRNAs (rasiRNAs), with a size of ~24 nt, are generated from both intergenic and repetitive genomic regions and are associated with DNA-methylation (Llave *et al.* 2002b, Mette *et al.* 2000, Hamilton *et al.* 2002, Baulcombe 2004). The biogenesis of hc-siRNA depends on transcription by RNA polymerase IV followed by dsRNA synthesis, which is catalyzed by the RNA-dependent RNA polymerase 2 (RDR2), and final processing into hc-siRNA duplexes by DCL3 (Herr *et al.* 2005, Xie *et al.* 2004, Kasschau *et al.* 2007). These hcsiRNA duplexes are incorporated into an AGO4 containing RISC guiding DNA- and histone-methylation (Zilberman *et al.* 2003).

The biogenesis of natural antisense transcript siRNAs (nat-siRNAs) primarily depends on the pairing of complementary natural antisense transcripts (NATs) (Borsani *et al.* 2005). The dsRNA is formed by complementary pairing of overlapping transcripts from opposing strands, hence they are called cis-nat-siRNAs opposing to the model of transnat-siRNAs generated by transcripts from different genomic loci with at least partial complementarity (Borsani *et al.* 2005, Axtell 2013). The dsRNA region from the NATs is processed into nat-siRNAs by DCL1 or DCL3 resulting in ~21 nt or ~24 nt sRNAs respectively (Zhang *et al.* 2012).

Secondary siRNAs (sec-siRNAs) are generated from a transcript, which is targeted by an sRNA resulting in the cleavage of the transcript and the recruitment of an RDR for the generation of an dsRNA that serves as a precursor for the secondary siRNA processing (Axtell 2013).

Trans-acting siRNAs (ta-siRNAs) are sec-siRNAs that are produced in phased pattern after an initial miRNA-triggered cleavage by DCL1 from a dsRNA that was transcribed from non-coding ta-siRNA coding loci (TAS) (Peragine *et al.* 2004, Vazquez *et al.* 2004). The initial non-coding TAS transcript is generated by Pol II, which is synthesized into a dsRNA after the initial cleavage by the RNA-dependent RNA polymerase 6 (RDR6) and stabilized by Suppressor of gene silencing 3 (SGS3) to prevent degradation (Yoshikawa *et al.* 2013). The phased generation of siRNAs through DCL4 processing results in siRNAs of equal length of 21 nt that are generated sequential head-to-tail, starting from the initial cleavage site (Vazquez *et al.* 2004, Fei *et al.* 2013). The phased ta-siRNA is assumed to function in a coordinated post-transcriptional regulation of multiple targets in *trans* (Coruh *et al.* 2014).

It was shown that 22 nt siRNAs are able to trigger sec-siRNA production at their target loci (Mlotshwa *et al.* 2008). The biogenesis of sec-siRNAs results in the spreading of siRNA production up- and downstream of the initial target locus, a phenomenon termed transitivity (Vaistij *et al.* 2002, Vasquez & Hohn 2013). The 21-22 nt siRNA triggered sec-siRNA pathway was shown to be involved in the silencing of epigenetically active transposable elements as well as in the virus-induced gene silencing (VIGS) (McCue *et al.* 2012, Nuthikattu *et al.* 2013).

1.3 Viral gene silencing suppressors

Plants developed RNA silencing mechanisms to defend against virus infections by detecting viral RNA and degrading them into sRNAs that are able to trigger a systemic silencing signal (Hamilton *et al.* 2002). This defense mechanism is known as virusinduced gene silencing (VIGS). Plant viruses coevolved gene silencing suppressors as a counter-defense to evade or suppress plant RNA silencing.

A multitude of mechanisms were uncovered for viral gene silencing suppressors acting in production, function, or stabilization of sRNAs (see Figure 3) (Peláez & Sanchez 2013, Pumplin & Voinnet 2013).



Figure 3: Model of the function of various viral gene silencing suppressors (yellow hexagons) on plants gene silencing components (Peláez & Sanchez 2013)

Examples for gene silencing suppressors impeding the production of siRNAs are the transactivator protein P6 from Cauliflower mosaic virus (*CaMV*) that interacts with DRB4, a cofactor of DCL4, and V2 protein of the *Tomato yellow leaf curl virus (TYLCV)* and P6 from *Rice yellow stunt virus (RYSV*), which both interact indirectly or directly with RDR6 respectively and thus block systemic RNA silencing via secondary sRNAs (Haas *et al.* 2008, Glick *et al.* 2008, Guo *et al.* 2013). The P0 protein from *Polerovirus*, the P38 protein from *Turnip crinkle virus (TCP*), and the P25 protein from *Potato virus X (PVX)* disrupt the effector complex by binding to it or promoting the degradation of the AGO1 and/or AGO2 proteins (Baumberger *et al.* 2007, Bortolamiol *et al.* 2007, Azevedo

et al. 2010, Chiu *et al.* 2010). The *Tombovirus* P19 protein competitively binds sRNAs and prevents their loading into AGO1, except for miR168, which is known to down-regulate AGO1 (Várallyay *et al.* 2010). The 126 kDa replicase subunit P126 from *Tobacco mosaic* virus (*TMV*) prevents the stabilization of sRNAs by interaction with HUA enhancer 1 (HEN1) (Vogler *et al.* 2007) and thus prevents sRNA being protected from degradation. The *Sweet potato chlorotic stunt virus* (*SPCSV*) is able to degrade siRNAs with a dsRNA-specific class 1 RNA endoribonuclease III (RNAse III) (Cuellar *et al.* 2009). The P1 protein from the P1/HC-Pro polyprotein from *WSMV* was shown to function as a suppressor of RNA silencing (Sentner 2008, Young *et al.* 2012), but the exact mechanism by which P1 acts is still unknown.

1.4 Objectives

Heterosis, the increased phenotypic performance of hybrids in comparison to their parental inbred lines, provided continuous improvements in stress resistance and yield to various crop plants. Although hybrid breeding is of high value for agriculture, the genetic and molecular mechanisms underlying this phenomenon are not yet understood, despite constant research for more than one century.

This study aims to analyze the contribution of small RNAs to the establishment of heterosis in maize.

It was shown for transgenic maize plants expressing P1/HC-Pro, a gene silencing suppressor from the *WSMV* that a perturbation of small RNAs in hybrids results in higher heterosis through an unknown mechanism (Thiemann 2011). This study aims to analyze the sRNA populations of induced vs. non-induced transgenic P1/HC-Pro plants by sRNA deep sequencing to investigate which sRNA pathways might be involved in heterosis formation in maize and elucidate the mode of action of P1/HC-Pro.

Many studies in various plant species uncovered differences between sRNA populations of inbred lines as well as their reciprocal hybrids. This study aims to identify heterosis-associated sRNAs by sequencing of sRNA populations of 21 maize inbred lines and association with heterosis for grain yield in their 98 resulting hybrids. The associated sRNAs will be characterized to elucidate their origin and putative mechanism of acting. The associated sRNAs will be tested for their value in the prediction of heterosis based on parental expression differences.

2. Material and Methods

2.1 Materials

2.1.1 Software

Table 1: Software used in this study

software

name	version	author/company	function	download source
Blast2GO	2.5.1	Conesa & Götz (2008)	functional sequence annotation	http://www.blast2go.com/ start-blast2go
BLASTn	2.2.26+	Camacho <i>et al.</i> (2009)	sequence alignment	ftp://ftp.ncbi.nlm.nih.gov/ blast/executables/blast+/ LATEST/
BLASTx	2.2.26+	Camacho <i>et al.</i> (2009)	sequence alignment	ftp://ftp.ncbi.nlm.nih.gov/ blast/executables/blast+/ LATEST/
Bowtie	0.12.9	Langmead <i>et al.</i> (2009)	sequence mapping	http://bowtie- bio.sourceforge.net/
Cufflinks	2.2.0	Trapnell <i>et al.</i> (2012)	transcriptome assembly, differential expression analysis	http://cufflinks.cbcb.umd.edu/
Circos	0.66	Krywinski <i>et al.</i> (2009)	data visuali- zation	http://circos.ca/software/ download
Java		Oracle Corp. Redwood City, CA USA	programming , language	http://www.java.com/
Java statis- tical class API (JSC)	1.0	Bertie (2005)	Java API, statistical analysis	http://jsc.nildram.co.uk/

software				
name	version	author/company	function	download source
MySQL		Oracle Corp. Redwood City, CA, USA	database system	http://www.mysql.com
PHP		The PHP Group	programming language	http://www.php.net
Picard Java API	1.67, 1.106		Java API; SAM/BAM file handling	http://picard.sourceforge.net/
psRNATarget		Dai & Zhao 2011	plant sRNA target predic- tion	http://plantgrn.noble.org/ psRNATarget/
R		Ihaka & Gentle- man (1996)	statistical analysis	http://www.r-project.org/
SAMtools	0.1.18	Li <i>et al. (</i> 2009)	SAM/BAM file handling	http://samtools.sourceforge.net
SRA toolkit	2.2.0	NCBI, Bethesda, MD, USA	SRA file handling	http://www.ncbi.nlm.nih.gov/ Traces/sra/?view=software

Table 1 (continued): Software used in this study

2.1.2 Plant material, growth and phenotyping

2.1.2.1 P1/HC-Pro gene silencing suppressor experiment

The experiment analyzing the effect sRNAs on the establishment of heterosis by alteration of the sRNA populations using a gene silencing repressor is based on transgenic plants containing an inducible P1/HC-Pro construct (see Figure 4), that was generated and provided by Dr. José Gutierrez-Marcos (University of Warwick, UK). The construct contains the gene BAR (bialaphos resistance) coding for a phosphinothricinacetyltransferase (PAT) providing a resistance to the herbicide Basta® (Bayer CropScience AG, Lyon, France) as a marker for transformed plants. The construct expresses the GAL4 transcription factor fused to a glucocorticoid receptor under the control of an UBI3 promoter. Under Dexamethasone-induction, the GAL4 transcription factor activates a bidirectional promoter resulting in transcription of the P1/HC-Pro gene silencing suppressor from the *wheat streak mosaic virus* (*WSMV*) and the β -glucoronidase (GUS) gene. GUS is used as a reporter for the expression of the transgene *in planta*.

The initial plants have a Hi-II background and were stably transformed using *Agrobacterium tumefaciens*. The plants were back-crossed for eight generations with the inbred line A188 to generate nearly homozygous transgenic inbred lines. The transgenic lines were selected by watering with 300 mg L⁻¹ of the herbicide Basta® (Bayer CropScience AG, Lyon, France). In addition to transgenic P1/HC-Pro plants with A188 background, the inbred lines A188 and H99 were included in the experiment for the generation of hemizygous inbred lines and hybrids. Backcrossing and selection were performed prior to this work.





The BAR gene under the control of the constitutive 35S-promotor serves as a herbicide selection marker for transformed plants. The Dex-inducible GR:GAL4-activator under the control of the UBI3-promotor activates the GAL4-promoter under Dex-induction and results in transcription of the gene silencing suppressor P1/HC-Pro from WSMV and GUS as transgene reporter.

In a randomized observer-blinded study, nucellus tissue of reciprocal transgenic and non-transgenic inbred lines (P1xA188 and A188xA188) or hybrids (P1xH99 and A188xH99) respectively were isolated one day after pollination (dap), transferred to plates containing modified Murashige-Skoog (MS) medium (Campenot *et al.* 1992) and cultured in the dark at 26 °C. The P1 expression was induced by adding 20 µmol L⁻¹ Dex to the medium. Half of the plates were grown by adding ethanol instead of Dex, to test for unintended growth effects. Two days after germination (dag) of the embryo (around 30 days after transferring the nucelli to the medium), the germ buds were transferred to standard MS medium (Murashige & Skoog 1962) and treated with Dex or ethanol as before. The plant height was determined every two days between 8 dag and 20 dag. After phenotyping, transgenic and non-transgenic progeny of the hemizygous transgenic parental plants were distinguished by GUS staining of leaf segments for the detection of the co-expressed GUS gene in transgenic plants. Three biological repli-

cates each of Dex-induced and non-induced transgenic hybrids, were flash-frozen in liquid nitrogen for later sRNA isolation. The tissue culture and phenotyping experiment was performed in advance of this work.

2.1.2.2 sRNA transcriptome analysis in inbred lines from a breeding population and hybrids of different heterosis-levels

For the sRNA/trait-association experiment, seedlings from 22 elite maize inbred lines and three hybrids were analyzed. 21 of the inbred lines (F037, F039, F043, F047, L024, L035, L043, P033, P040, P046, P048, P063, P066, S028, S036, S044, S046, S049, S050, S058, S067) and the three hybrids (P033xF047, S028xF039, S027xL024) were obtained from a 14x7 half diallel factorial mating scheme of Dent and Flint maize of the breeding program of the University of Hohenheim (Germany). The seven Flint lines constisted of four inbred lines with European Flint background (F037, F039, F043, F047) and three with Flint/Lancaster background (L028, L035, L043). The 14 Dent lines comprise six lines with lodent background (P033, P040, P046, P048, P063, P066) and eight lines with an Iowa Stiff Stalk Synthetic background (S028, S036, S044, S046, S049, S050, S058, S067). Three hybrids were chosen to cover low (S028xF047), intermediate (S028xF039) and a high (P033xF047) mid-parent heterosis (MPH) for grain yield (GY). In addition the inbred line B73 was included for a direct comparison to the reference genome.

All lines were grown under controlled conditions (25 °C, 16 h day, 8 h night, 70 % air humidity) for seven days and flash-frozen in liquid nitrogen. Five individuals of the same genotype were pooled before sRNA isolation to reduce variability.

Phenotypic data of the 21 inbred lines and three hybrids from the breeding program of the University of Hohenheim were collected from field trials, the inbred lines in 2003 and 2004 at five locations and the hybrids in 2002 at six locations in Germany (Schrag *et al.* 2006). GY field data were measured in Mg ha⁻¹ adjusted to 155 g kg⁻¹ grain moisture. All 98 hybrids show positive MPH and GY levels higher than both their inbred parents. Thus, best-parent heterosis (BPH), the trait-specific performance of the hybrid relative to the best performing parent, is also positive for all hybrids.

2.2 Molecularbiological methods

2.2.1 RNA isolation

Total RNA isolation was performed using the mirVana miRNA Isolation Kit (Life Technologies Corp., Carlsbad, CA, USA). The quality of the isolated RNA was confirmed by photometrical and gelelectrophoretic analyses prior to sequencing.

2.2.2 Small RNA sequencing

sRNAs library preparation was performed from total RNA using the TruSeq SBS Kit v5 (Illumina Inc., San Diego, CA, USA) by the sequencing service of either Eurofins MWG GmbH (Ebersberg, Germany) or LGC Genomics GmbH (Berlin, Germany). All sRNA libraries were indexed with barcodes and sequenced on an Illumina Hi-Seq 2000 (Illumina Inc. San Diego, CA, USA) with up to four samples per sequencing lane.

2.3 Computational methods

The microarray re-annotation was performed using custom PHP scripts. All other analyses were performed with custom Java-programs if not stated differently in the text. Statistical analyses in some of the custom Java-programs were performed using the Java Statistical Classes (JSC) API (Bertie 2004).

The microarray re-annotation was performed on a Dell Optiplex 980 (Dell Inc., Round Rock, TX, USA) with one Intel i5 2.67 GHz quad-core CPU (Intel Corp., Santa Clara, CA, USA) and 16 GB RAM with the two operating systems Windows 7 Professional (Microsoft Corp., Redmond, CA, USA) and Debian Linux (Debian Project; Software in the Public Interest Inc., New York, NY, USA).

The analysis of the sRNA sequencing data was performed on a custom workstation with one Intel Xeon E5-2620 2 GHz hexa-core CPU (Intel Corp., Santa Clara, CA, USA) and 32 GB RAM and a custom workstation equipped with two AMD Opteron 6272 2.1 GHz 16-core CPUs (Advanced Micro Devices Inc., Sunnyvale, CA, USA) and 64 GB RAM. Both workstations were running with Debian Linux (Debian Linux (Debian Project; Software in the Public Interest Inc., New York, NY, USA).

2.3.1 Re-annotation of the 46k-maizearray

2.3.1.1 Oligonucleotides localization and target gene identification

The oligonucleotide sequences of the 57k maizearray (Gardiner et al. 2005) that comprises all oligonucleotides of the 46k-maizearray were aligned to the B73 maize referdownloaded ence genome (RefGen_v2; from http://ftp.maizesequence.org/current/assembly/, April 2012) using BLASTn (standalone BLAST, version 2.2.26+; Camacho et al. 2009) with a maximum e-value of 0.0001 and a word-size of 20 for the identification of their potential targets. Alignments with more than three mismatches, insertions or deletions (indels) were rejected to reduce ambiguity of the results. All full length matches of oligonucleotides were analyzed for sense and antisense strand of exon or intron annotations of the B73 working gene set (WGS version 5a.59, downloaded from: http://www.maizesequence.org/current/assembly/) and repeat annotations from the TE Consortium (version 5a, downloaded from: http://ftp.maizesequence.org/current/repeats/). Alignments of oligonucleotide fragments were tested for the residual fragment within 20 kbp. This sequence interval was chosen to equal the maximum length of maize intron sequences (Schnable et al. 2009). The sum of the two located fragment lengths needed to exceed the full length oligonucleotide minus 10 bp to be accepted as an oligonucleotide overlapping a putative splice site. If no second fragment was found and the mapped fragments length exceeds the oligonucleotide length minus the blast word-size of 20 and additional three mismatches/indels, the oligonucleotide was accepted as putatively overlapping a splice site. In this case, a second fragment might have been missed due to the BLASTn word-size. All maizearray oligonucleotides putatively overlapping a splice site were mapped to maize WGS cDNA sequences (version 5a.59) using BLASTn with identical parameters as above.

The number of genes, transcripts or repetitive elements covered by each oligonucleotide was identified. Furthermore, the number of oligonucleotides representing a gene or repeat was counted.





Figure 5: maizearray re-annotation procedure

2.3.1.2 Functional annotation of target genes

The functional annotation of the identified target genes of the oligonucleotide sequences was generated using Blast2GO (version 2.5.1; Conesa & Götz 2008). The sequences were aligned to the NCBI non-redundant protein sequences (nr) database using the BLASTx (Altschul *et al.* 1997) routine with maximum e-value set to 0.01 and maximum number of hits to 20. GO-terms were collected for all successfully aligned oligonucleo-tides.

2.3.2 Sequencing data processing

2.3.2.1 sRNA sequencing data processing

sRNA sequencing datasets of the P1/HC-Pro transgenic hybrids and the 21 inbred lines and three hybrids from the breeding program of the University of Hohenheim as well as the inbred line B73 were obtained in FastQ file format from the sequencing service Eurofins MWG GmbH (Ebersberg, Germany) or LGC Genomics GmbH (Berlin, Germany) respectively. The sequences were already trimmed from 5'-adapter sequences and allocated to sequencing libraries by their sequencing barcode.

The RNAseq and sRNAseq sequence data from five day old shoots from the study by Regulski *et al.* (2013) were used for the sRNA transcriptome interaction study. The datasets were downloaded in sequence read archive (SRA) file format from NCBI GEO (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39232) and converted into FastQ file format using the program fastq-dump (NCBI SRA Toolkit, version 2.2.0).

The sequence data were both trimmed from the Illumina 3'-adapter sequence and from sequence regions with sequencing quality below 99.9 % (Phred score <30). All sequences in the range of 15 nt to 40 nt were retained. Redundant sRNA sequences were merged using a custom R-script to obtain raw read counts for each sequencing library.

2.3.2.2 sRNA sequencing data normalization

The raw read counts from transgenic hybrid lines were normalized to 1 M read counts per sequencing library (rpm) to enable direct comparison of the replicates' expression values by averaging 1000 random sampling runs of 1 M sequences each from the raw sequencing data. A MySQL database was generated for further analyses containing both raw and normalized expression data.

The raw read counts from the 22 inbred lines and three hybrids for the association study were integrated in an expression dataset of the dimensions g×n with g being the genotypes (22 inbred lines and three hybrids) and n the number of distinct sRNA sequences over all samples. The expression dataset was quantile normalized according to the method by Bolstad *et al.* (2003) with a modification preventing the allocation of normalized read counts samples without any expression in the raw expression dataset, resulting in quantile normalized read counts per library. To allow for direct comparability of sequencing libraries with different sequencing depths, the quantile normalized read counts were scaled to one million reads per library, resulting in read counts per million quantile normalized (rpmqn) reads. A MySQL database was generated for further analyses containing both raw and normalized expression data.

The raw sRNA read counts from two inbred lines, sequenced in triplicates, from the Regulski *et al.* (2013) dataset were quantile normalized (Bolstad *et al.* 2003) with the same modification as stated before and scaled to one million reads per library (rpmqn).

2.3.2.3 Transcriptome sequencing data processing

The raw read counts from the transcriptome sequencing project of five day old seedlings from two maize inbred lines, sequenced in triplicates, from the Regulski *et al.* (2013) dataset were trimmed from the Illumina 3'-adapter sequence and from sequence regions with sequencing quality below 99.9 % (Phred score <30).

2.3.3 P1/HC-Pro gene silencing suppressor experiment

2.3.3.1 Calculation of best parent heterosis increase

The increase of best-parent heterosis (BPH) for the induced vs. non-induced transgenic P1/HC-Pro hybrids was calculated for the mean of the BPH values. The significance of the increase of heterosis (BPH for growth rate) in induced transgenic P1/HC-Pro plants was analyzed by a one-sided Student's t-test and confirmed by a permutation test of the BPH values from all possible hybrid/inbred combinations with 1 M resampling runs.

The growth rate data for inbred lines and hybrids of both induced and non-induced transgenic P1/HC-Pro plants are shown in Appendix Table 3.

2.3.3.2 sRNA population complexity analysis

The sRNA population complexity, defined as the number of distinct sRNA sequences in a set of sequences of a certain sample size, was calculated separately for each individual hybrid sample via bootstrap analysis with replacement with 1000 sampling runs for the sample sizes: 100 k, 200 k, 500 k, 1 M, 1.5 M, 2 M, 2.5 M, 3 M, 3.5 M, 4 M, 4.5 M, and 5 M sequences. The significance of the sRNA population complexity reduction in

the induced transgenic plants in comparison to the non-transgenic plants was analyzed in one-sided heteroscedastic Student's t-tests and adjusted for FDR of 10 % (Benjamini & Hochberg 1995).

2.3.3.3 sRNA length distribution analysis

To analyze effects of P1/HC-Pro on specific sRNA lengths, the normalized read counts were summed up by sequence length separately for all individual hybrid samples to calculate the fraction of reads of a specific length of all sRNA sequences. The fractions for read counts of the sRNA lengths from 18 nt to 28 nt of transgenic and non-transgenic plants were tested for differences in two-sided heteroscedastic Student's t-tests and adjusted for FDR of 5 % (Benjamini & Hochberg 1995).

2.3.3.4 Analysis for differentially expressed miRNAs

For the identification of differentially expressed miRNA, the distinct sRNA sequences of all hybrid samples were mapped to known maize pre-miRNA precursor sequences from miRBase (release 20; Griffiths-Jones *et al.* 2006). All sequences mapping to precursors were tested for differential expression by two-sided heteroscedastic Student's t-tests and adjusted for FDR of 5 % (Benjamini & Hochberg 1995).

2.3.4 sRNA transcriptome analysis in inbred lines from a breeding population and hybrids of different heterosis-levels

2.3.4.1 Sequence mapping of sRNA sequences

The distinct sRNA sequences of the 22 inbred lines and three hybrids were mapped without mismatches to the maize B73 reference genome (RefGen_v2; downloaded from http://ftp.maizesequence.org/current/assembly/, April 2012) to identify their putative loci of generation/effect, using the short read aligner Bowtie (version 0.12.9; Langmead *et al.* 2009) and exported in Sequence Alignment/Map format (SAM) file. Sorted and indexed binary Sequence Alignment/Map format (BAM) files were generated from the SAM file using the SAMtools (version 0.1.18, Li *et al.* 2009) for faster annotation of the sRNA data. All custom Java-programs generated for read annotation that are parsing
either SAM or BAM files are using the Picard Java API (version 1.67, downloaded from: http://picard.sourceforge.net).

For each distinct sRNA, the number of mapping positions to the B73 reference genome was determined to identify their degree of conservation. The fraction of sRNAs that are unmapped, uniquely mapping, or mapping 2-10 times, 11-100 times, or >100 times to the reference genome were determined. The mapping position counts were tested in two-sided Student's t-tests for differences between Dent and Flint inbred lines as well as inbred lines and hybrids.

The distinct sRNA sequences were mapped to the precursor sequences of all known maize miRNAs from miRBase release 20 (Griffiths-Jones *et al.* 2006).

2.3.4.2 Annotation of sRNA sequences

A local B73 annotation MySQL database was generated based on gene annotations from the B73 working gene set (WGS version 5a.59, downloaded from: http://www.maizesequence.org/current/assembly/) and TE Consortium repeat annotations (version 5a, downloaded from: http://ftp.maizesequence.org/current/repeats/), which were filtered for redundant entries. All regions neither covered by gene nor repeat annotations were defined as intergenic. The genome coverage of the annotation types was calculated for sequential windows of 1 Mbp width.

The mapped sRNAs were annotated in a relational database using the B73 annotation database, allowing the identification of loci of generation/effect based on annotation.

2.3.4.3 sRNA differential expression analysis

The sRNA differential expression state x_s for the sRNA s in the comparison of the two inbred lines i and j with read count c_i and c_j respectively and arbitrary parameters defining the threshold for minimal read count c_{min} stating an expressed sRNA and for the minimal expression fold-change f_c for differential expression between c_i and c_j is defined as follows:

$$x_{s} = \begin{cases} 1 & if \quad c_{l} \ge c_{min} \land c_{h} \ge c_{l} \cdot f_{c} \\ 1 & if \quad c_{l} < c_{min} \land c_{h} \ge c_{min} \cdot f_{c} \\ 0 & else \end{cases}$$
 Formula 1: Differential expression calculation

with

$$c_h = max(c_i, c_j)$$
 Formula 2: High parent expression value

and

$$c_l = min(c_i, c_j)$$
 Formula 3: Low parent expression value

An sRNA is defined as differentially expressed between the two inbred lines i and j if $x_s=1$. Thus, the sRNA is either differentially expressed if

• the lower expressed inbred lines read count c_l equals or exceeds the minimal read count c_{min} and the higher expressed inbred lines read count c_h exhibits an expression fold-change equal or higher than f_c relative to c_l

or

 the lower expressed inbred lines read count c_l is below the expression threshold and the higher expressed inbred lines read count c_h exhibits an expression foldchange equal or higher than f_c relative to the expression threshold c_{min}. 2.3.4.4 sRNA population differential expression distance calculation

The (qualitative) binary distance D_b of two inbred lines i and j, based on differential expression of the sRNAs in an sRNA population, with n_s defining the number of distinct sRNA sequences in the sRNA population is calculated as follows:

$$D_b(i,j) = \sqrt{\frac{1}{n_s} \sum_{s=1}^{n_s} x_s}$$
Formula 4: Binary distance calculation for a pair of inbred lines

The binary distance D_b gives a measure of the fraction of differentially expressed sRNAs in the sRNA population between the two inbred lines. The value of D_b ranges from 0 if the two populations do not exhibit differentially expressed sRNAs to 1 in case of all sRNAs of the sRNA populations being differentially expressed between the two inbred lines.

The (quantitative) euclidean distance D_e of two inbred lines i and j is based on their expression values of the sRNAs in an sRNA population of n_s distinct sRNA sequences. The expression of a specific sRNA s of the two inbred lines is given by $c_i(s)$ and $c_j(s)$. The euclidean distance is calculated as follows:

$$D_e(i,j) = \sqrt{\frac{1}{n_s} \sum_{s=1}^{n_s} (c_i(s) - c_j(s))^2}$$

Formula 5: Euclidean distance calculation for a pair of inbred lines

2.3.4.5 sRNA population diversity analysis

The sRNA population diversity was calculated to reveal the number of sRNAs present in two lines, thus potentially conserved, and sRNAs present in only one of the two lines. The sRNA population diversity was calculated based on sRNA sequences with an expression fulfilling or exceeding the expression threshold $c_{min}=0.5$ rpmqn for at least one inbred line in a comparison of two inbred lines, or at least one inbred line or the hybrid for the inbred-hybrid triplets. The population diversity was analyzed separately for all possible 14x7 inbred combinations between the two heterotic groups (Dent and Flint) as well as for all inbred-hybrid triplets. The average population diversity was calculated for all 98 inbred combinations.

2.3.4.6 Inbred line differentially expressed sRNA population diversity comparison and grouping

The sRNA population diversity based on differentially expressed sRNAs was determined for all possible inbred line pairs from the 21 elite maize inbred lines from the University of Hohenheim. This analysis was performed to test if distance based on differential expression somehow reflects the genetic distances of those lines.

The sRNA differential expression x_s (Formula 1) was calculated with the arbitrary parameters c_{min} =0.1 rpmqn and f_c =100 for all distinct sRNAs with a length of 18 nt to 28 nt.

The diversity is calculated as the binary distance D_b (Formula 4) for all distinct sRNAs of the two inbred lines i and j. n_s is defined as the total number of distinct sRNAs of line i that are differentially expressed compared to any of the other 20 inbred lines, consecutively defined as line j. Note that distinct sRNAs are defined based on their sRNA sequences, not their different expression patterns.

The grouping of inbred lines is performed by the first two components of a principal component analysis (PCA) based on the binary distance D_b of sRNA differential expression. The PCA was performed using a custom R-script including the library FactoMineR (version 1.26, Husson *et al.* 2007). The PCA plot for the first two components is analyzed for inbred line grouping based on D_b .

2.3.4.7 Association of inbred parent sRNA expression with hybrid trait-values

The association of parental differential sRNA expression patterns to hybrid trait values, in this study mid-parent heterosis for grain yield (MPH for GY), was performed according to the method established for the association of transcriptome data by Frisch *et al.* (2010). Thus, the hybrids of the factorial mating scheme were sorted by their ascending hybrid trait values, which is MPH for GY, and grouped into equally sized classes, resulting in one group with low (L) and one group with high (H) MPH for GY hybrids respectively. The method by Frisch *et al.* (2010) considers only positively associated transcripts, thus the association method in this study was extended to include the identification of negatively trait associated sRNAs.

For each distinct sRNA in the range of 18 to 28 nt of the 21 inbred lines, the association is performed by counting the occurrence of sRNA differential expression between the

hybrids inbred parents for the two groups L and H as o_L and o_H respectively. The probability P_s of the sRNA s being associated to MPH for GY is estimated via the binomial distribution probability function, depending on o_L and o_H under the condition that sRNA differential expression occurs with equal probability (p = 0.5) in the two groups L and H as follows:

$$P_s = \sum_{k=k_{min}}^{k_{max}} Bin_{n,p}(k)$$
 Formula 6: Association probability calculation

with

$$Bin_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}$$
 Formula 7: Binomial probability mass function and

 $n = (o_H + o_L)$ Formula 8: Sum of differentially expressed sRNAs in both groups

and in case of positive association:

$$k_{min} = 0, k_{max} = o_L \quad if \quad o_L > o_H$$

or in case of negative association:

$$k_{min} = o_H, k_{max} = (o_H + o_L)$$
 if $o_L \le o_H$
ters for negative association

respectively.

The p-values are finally adjusted for FDR of 5 % correction (Benjamini & Hochberg 1995) for P_s <0.05. All sRNAs succeeding this correction are considered as heterosis-associated sRNAs (ha-sRNAs).

To control for associations derived from random noise in the data, a permutation analysis with 100 runs was performed by either randomly re-assigned hybrid trait values or inbred line labels, resulting in re-grouping of the classes L and H or differential expression patterns respectively. 2.3.4.8 Correlation analysis of sRNA-based distance of inbred parents and heterosis

Different distance measures of differentially expressed sRNAs as well as associated sRNAs were tested for correlation with trait values to reveal putative inherent links.

A correlation of the binary and euclidean distances (D_b and D_e) of all differentially expressed sRNAs in at least one of the 14x7 inbred combinations with the hybrid trait MPH for GY was performed.

For the ha-sRNAs, correlation analyses were performed separately for the binary distances of the positively and negatively associated sRNAs (as $D_{b,pos}(i,j)$ and $D_{b,neg}(i,j)$ respectively) between the inbred parents of the 98 hybrids.

A combined binary distance $D_{b,com}(i,j)$ joins the binary distances of the positively $D_{b,pos}(i,j)$ and negatively associated sRNAs, weighted by the number of distinct sRNAs in the results of positively n_{pos} and negatively n_{neg} associated sRNAs for the two inbred parents i and j as follows:

$$D_{b,com}(i,j) = \frac{(D_{b,pos}(i,j) \cdot n_{pos}) + ((1 - D_{b,neg}(i,j)) \cdot n_{neg})}{n_{pos} + n_{neg}}$$

Formula 11: Combined binary distance of pos. and neg. associated sRNAs

 $D_{b,com}(i,j)$ was calculated analogously to $D_{b,pos}(i,j)$ and $D_{b,neg}(i,j)$ for all inbred parents of the 98 hybrids.

2.3.4.9 sRNAome-based prediction of hybrid trait values

The prediction of hybrid trait values for unknown hybrids based on sRNA expression data of inbred lines using a linear regression model is performed analogously to Frisch *et al.* (2010) and Fu *et al.* (2012) using a linear regression model with an extension including the neg. ha-sRNAs.

The accuracy of the prediction of MPH for GY was performed with the procedures for type 0 and type 2 prediction as described in Fu *et al.* (2012) with equal parameters for the identification of ha-sRNAs in the estimation datasets as applied for the association on the full dataset (see 2.3.4.7).

2.3.4.10 ha-sRNA length distribution enrichment analysis

The number of distinct sRNAs was summed up separately for positive and negative hasRNAs as well as bootstrap samples with equal numbers of distinct sRNAs as the positive or negative ha-sRNA sets respectively. The bootstrap analysis was performed with replacement from all distinct sRNAs of the 21 inbred lines with 1000 runs for the retrieval of enrichment probabilities.

2.3.4.11 ha-sRNA annotation distribution analysis

The characterization loci of ha-sRNAs generation or is performed by retrieving the annotation of each distinct ha-sRNA, defined by all loci it could be mapped to on the B73 reference genome. Each ha-sRNA was accounted for an annotation type (gene, repeat, intergenic) or, if multiple annotation types were identified, it was accounted for intersections of multiple annotation types.

2.3.4.12 Distribution of ha-sRNAs at and around gene loci

The distribution of ha-sRNAs at and around gene loci was analyzed by parsing the hasRNA mapping data (BAM file) for sRNAs that are located 5 kbp up-/downstream of gene annotations or in the exons/introns of genes. The 5 kbp regions up-/downstream of the genes were analyzed in 250 bp windows. The exon/intron regions were analyzed relative to the length of the annotation in 0.5 % windows.

2.3.4.13 Enrichment analysis for ha-sRNAs for repeat super-families and families

The enrichment and depletion probability of ha-sRNA lengths for repeat super-families and/or families were tested by bootstrap analysis with 1000 runs. The analysis was per-formed for all, and additionally separately for the positive and the negative ha-sRNAs. The random bootstrap samples were of equal length distribution and sample size as the specific reference sample. The analysis considered only sRNAs that were attributed to one repeat super-family/family to exclude ambiguity.

2.3.4.14 Genome-wide enrichment analysis for ha-sRNAs

All maize chromosomes were tested for enrichment/depletion of positive or negative hasRNAs in 1 Mbp windows by bootstrap analysis with 1000 runs to reveal loci of generation/effect are being present at specific genome regions. The random bootstrap samples were of equal size and length distribution as the reference set made of the positive or negative ha-sRNAs respectively.

2.3.4.15 ha-sRNA expression pattern analysis

The dominance to additivity (d/a) expression pattern calculation was performed as described in Li *et al.* (2012) for all differentially expressed sRNAs (parameters similar to association analysis: c_{min} =0.5 and f_c =2) and separately for both the subsets of negative-ly and positively ha-sRNAs for the three sequenced hybrids.

The ratios of differentially expressed ha-sRNAs, separately for all sRNA lengths, belonging to either presence-absence or fold-change pattern were calculated for all 14x7 inbred parent combinations separately for all sequence lengths to identify predominant expression patterns. A presence-absence expression pattern is defined as the low expressed parent's expression lying below the threshold of 1 rpmqn.

2.3.4.16 ha-sRNA target prediction

The target transcript identification of ha-sRNAs, putatively acting either via posttranscriptional gene silencing or translational inhibition, was performed using psRNATarget (http://plantgrn.noble.org/psRNATarget/; Dai & Zhao 2011) using the transcript database "Zea mays (maize), cds, PlangGDB genomic project," for all hasRNAs and separately for each ha-sRNA length with standard parameters. The parameter hspsize was set to 18 or 19 respectively for the 18 nt and 19 nt ha-sRNAs to allow a target prediction.

2.3.5 Analysis of effects on the transcriptome mediated by ha-sRNAs

2.3.5.1 Correlation of ha-sRNA expression and microarray transcriptome expression data

To identify putative target transcripts that are regulated by ha-sRNAs in linear dependent mechanism, the sRNA expression data of all ha-sRNAs of the 21 inbred lines was correlated with transcriptome expression (see 2.3.4.16). The transcriptome expression data was generated from the identical biological material in a previous study (Thiemann *et al.* 2010) using the 46k maize oligonucleotide array (Gardiner *et al.* 2005). Correlations were performed for all ha-sRNA/transcript pairs where the ha-sRNA maps at or adjacent to a gene locus of a transcript on the B73 reference genome, covered by the 46k maizearray as identified by the re-annotation (see 2.3.1.1). The p-values of correlated ha-sRNA/transcript were adjusted for multiple testing via FDR correction (Benjamini & Hochberg 1995) for p<0.05.

The ha-sRNAs showing significant correlation to target transcriptome expression data were characterized by their length distribution.

2.3.5.2 sRNA differential expression analysis

ha-sRNAs from the sRNA-association study were extracted from the sRNA sequencing dataset generated by Regulski *et al.* (2013) and tested for differential expression of the average of the replicates with equal parameters as described in chapter 2.3.4.3.

2.3.5.3 RNAseq differential expression analysis

To obtain transcript expression values from the RNAseq dataset by Regulski *et al.* (2013), the processed RNAseq reads were aligned to the B73 reference genome (RefGen_v2) using Tophat (version v2.0.11, Trapnell *et al.* 2009). The aligned reads were assembled using Cufflinks (version v2.2.0, Trapnell *et al.* 2012) and annotated by the B73 working gene set (version 5a.59). All replicates were merged using Cuffmerge (version v1.0.0, Trapnell *et al.* 2012). Differentially expressed genes/transcripts were identified using Cuffdiff (version v2.2.0, Trapnell *et al.* 2012).

2.3.5.4 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions

The ha-sRNAs mapping to differentially expressed genes or 1 kbp up- or downstream were identified and tested for enrichment in a bootstrap analysis with equal sRNA read length distribution in 1000 runs.

3. Results

3.1 Molecularbiological analyses

3.1.1 Small RNA sequencing

The sequencing of induced and non-induced transgenic P1/HC-Pro plants resulted in a total of 97.7 M raw 50 bp sequence reads, with the individual replicates ranging from 13.62 M to 23.3 M raw sRNA sequencing reads.

The sequencing of the 22 inbred lines and 3 hybrids resulted in total in 504.36 M raw 50 bp sequence reads. The individual sequence libraries for the samples exhibit a range from 12.85 M to 37.31 M raw reads.

3.2 Computational analyses

3.2.1 Re-annotation of the 46k-maizearray

The re-annotation of the 46k-maizearray was published and can be accessed via: http://www.maydica.org/suppl/57_49__S_table1.xlsx

3.2.1.1 Oligonucleotide localization and target gene identification

The BLASTn-search for alignments of oligonucleotides of the 57k-maizearray on the maize genome resulted in 3,376,312 alignments. 584,589 of these alignments were of full length, covering 37,351 (65.01 %) of the 57k oligonucleotides. 1,339,243 alignments were aligned containing up to three indels or mismatches corresponding to 48,756 (84.86 %) of the 57k-maizearray oligonucleotides. 214,386 fragmented or partial alignments were filtered as potentially overlapping splice sites for 10,873 (18.93 %) distinct oligonucleotides.

The annotation of the full length alignments with the maize WGS 5a.59 dataset resulted in 40,975 exons covered by 30,733 oligonucleotides and 13,014 intronic sequences covered by 5,405 oligonucleotides. The mapping of fragmented or partial alignments to B73 WGS cDNAs resulted in 7,125 transcripts represented by 5,355 oligonucleotides spanning intronic regions. The annotation of oligonucleotides to the antisense strand of annotations resulted in 5,103 exons that were represented by 5,101 oligonucleotides and 8,049 introns covered by 2,099 maizearray oligonucleotides respectively. 3,306 repeat annotations from the TE Consortium were covered by 1,190 oligonucleotide probes in sense and 1,796 annotations by 673 oligonucleotides on the antisense strand respectively.

In total 32,210 (73.98%) of the 46k maizearray oligonucleotides were annotated to at least one known gene or repeat, including annotations of introns as well as oligonucleotides oriented to the antisense strand of genes. 29,861 (68.59%) of the 46k maizearray probes represented at least one gene (exon, spliced exon, repeat; sum of the last three rows in Table 2). 23,369 (53.68%) of the 46k oligonucleotides accounted for the expression of a single gene (repeat; sum of the last thwo rows in Table 2). 17,241 (30.01%) of the 46k maizearray oligonucleotides could be furthermore assigned to a single splice-form of the gene. The results are summarized in Table 2 for both maizearray platforms.

	57k maizearray	46k maizearray
unannotated	16,760 (29.18 %)	11,326 (26.02 %)
antisense/intronic	3,811 (6.63 %)	2,349 (5.39 %)
multiple genes	8,473 (14.74 %)	6,492 (14.91 %)
single gene, multiple transcripts	11,167 (19.44 %)	9,715 (23.32 %)
single transcript	17,241 (30.01 %)	13,654 (31.36 %)

Table 2: Total and relative re-annotation results for the maize oligonucleotide array platforms

3.2.1.2 Functional annotation of target genes

The functional annotation of the maize oligonucleotide array resulted in 47,562 target genes covered by 32,745 (57 %) of the microarray probes. In total 38,144 GO-terms, collated to 30,546 (53.17 %) transcript-annotated (exon/spliced exon) oligonucleotides, were obtained, resulting in twice as many GO-annotated oligonucleotides.

3.2.2 Sequencing data processing

The raw sequencing data and raw as well as normalized read counts of processed sRNA reads of both experiments were published on NCBI GEO under the accession GSE51662 and can be accessed via:

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE51662

3.2.2.1 sRNA sequencing data processing

After 3'-adapter and quality trimming of the raw HC-Pro hybrid sequencing data, a total of 42.5 M processed sequences with a length of 15 nt to 40 nt were retained. The number of processed sRNA reads varied in the replicates from 5.38 M to 12.4 M. The redundant sequences of all six samples from the inducible transgenic P1/HC-Pro hybrids allocated to 12.8 M distinct sRNA reads, with the replicates ranging from 1.16 M to 3.25 M distinct reads (see Appendix Table 4).

After 3'-adapter and quality trimming of the sRNA sequencing data of the reference line B73 and the 21 inbred lines and three hybrids from the breeding program of the University of Hohenheim, 380.93 M processed reads (75.53 % of the raw reads) with a length of 15 nt to 40 nt were retained. The number of processed reads (15 nt to 40 nt) ranged from 8.95 M to 31.14 M for the individual sequence libraries (see Figure 8). The length distribution of the processed sRNA sequences exhibited a major peak at 24 nt and a minor peak around 21 nt, which in some lines was part of a plateau of sRNAs ranging from 21 nt to 23 nt with nearly equal amounts (see Figure 6).



Figure 6: sRNA length distribution of raw total read counts of the reference inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of Hohenheim After merging redundant sequences, in total 23.31 M distinct sRNA sequences were identified for the 22 inbred lines and three hybrids. The number of distinct sRNAs for the samples was between the limits of 1.34 M to 4.29 M. The length distribution of the distinct reads showed a major peak at 24 nt for all sequence libraries (see Figure 7) (see Appendix Table 5).



Figure 7: sRNA length distribution of distinct read counts of the reference inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of Hohenheim

3.2.2.2 sRNA sequencing data normalization

The varying sequencing depths of B73, 21 inbred lines and three hybrids from the breeding program of the University of Hohenheim, with raw sRNA read counts ranging from 8.95 M to 31.14 M, were normalized to allow for direct comparison of expression

data. The normalization resulted in normalized sRNA sequence libraries with a total of ~15 M quantile normalized read counts (see Figure 8, see Appendix Table 6).

Figure 8: Total sRNA read counts before (raw) and after quantile normalization of the reference inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of Hohenheim



The length distribution of the quantile normalized reads exhibited a major peak at 24 nt and a minor plateau at 21 to 22 nt for the sequence libraries of all inbred lines and hybrids (see Figure 9). The variation between the raw sequence libraries was reduced by the quantile normalization (see Figure 7 and Figure 9 in comparison).



3.2.3 P1/HC-Pro gene silencing suppressor experiment

3.2.3.1 Calculation of best-parent heterosis increase

The significance of best-parent heterosis (BPH) was confirmed (p<0.05) by the t-test as

well as the permutation test. The average increase of BPH for growth rate was 15.29 % for the induced relative to the non-induced transgenic P1/HC-Pro hybrids (see Figure 10).

Figure 10: BPH for growth for induced/non-induced transgenic P1/HC-Pro hybrid plants

* significant difference p<0.05 (one-sided t-test and permutation test)



3.2.3.2 sRNA population complexity analysis

The bootstrap analysis of the sRNA population complexity of induced and non-induced transgenic P1/HC-Pro plants revealed a significant (p<0.05) reduction of the population complexity by nearly 30 % (see Figure 11 and Appendix Table 7).



3.2.3.3 sRNA length distribution analysis



There was a non-significant reduction of 22 and 24 nt sRNAs in the induced transgenic P1/HC-Pro plants. The decrease of 24 nt sRNAs in the induced hybrids and high variance can be attributed to one of the three replicates deviating from the others (see replicate 1 of the induced transgenic P1/HC-Pro hybrids in Appendix Table 8).

3.2.3.4 Analysis for differentially expressed miRNAs

The mapping of the sequence reads to known maize pre-miRNA precursors from miRBase (release 20, Griffiths-Jones *et al.* 2006) did not reveal any differential expressed miRNAs between the induced and non-induced hybrids.

3.2.4 sRNA/trait-association experiment

3.2.4.1 Sequence mapping of sRNA sequences

The sequence mapping of sRNAs of the inbred lines and the hybrids to the B73 reference genome resulted in 82.03 % mapping sRNAs for the inbred line B73 and fractions in the range of 44.85 % to 58.10 % for the 21 maize inbred lines from the University of Hohenheim. The three hybrids exhibited a higher fraction of mapped sRNAs (59.87 % to 65.28 %) in comparison to the inbred lines (see Figure 13 and Appendix Table 9). This difference was found to be significant (p=0.0113) (see Appendix Table 10). The inbred lines of the two heterotic groups (Flint/Dent) did not show a significant difference in the fraction of mapping sRNAs. The mapping counts of sRNAs to the B73 reference genome showed the highest fraction of uniquely mapping sRNAs for the inbred line B73 (64.6 %). The 21 inbred lines ranged from 29.58 % to 42.06%. The hybrids in comparison to the inbred lines (except for B73) showed a significantly (p=0.00548) higher fraction (41.41 % to 44.15 %) of uniquely mapping sRNAs (see Figure 13 and Appendix Table 9).



Figure 13: sRNA mapping count distribution of the reference inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of Hohenheim

For all mapping count groups with multiple mapping positions (> 2 mapping positions), B73 did neither deviate from the other 21 inbred lines. Nor did the inbred lines of the two heterotic groups differ from each other. In contrast the 21 inbred lines deviated from the three hybrids (p<0.01) except for the mapping group containing sRNAs mapping 11-100 times (see Figure 13 and Appendix Table 10).

3.2.4.2 sRNA population diversity analysis

The sRNA population diversity analysis for the 14x7 inbred pairs of the two heterotic groups resulted on average in nearly equal fractions of sRNAs specific to either inbred lines of the Dent or Flint heterotic groups and slightly higher overlap of both parents. The fraction of sRNAs specific for Dent lines had a mean of 32.73 % \pm 2.37 %. Nearly identical the Flint lines had a mean of 31.69 % \pm 2.4 %. The overlapping fraction of sRNA populations between the inbred parents of the heterotic pools, with a mean of 35.58% \pm 4.19%, exhibited a higher variation in comparison to the parent-specific fractions (see Figure 14 and Appendix Table 11).



Figure 14: Average inbred parent sRNA population diversity analysis from 98 inbred line pairs of two heterotic groups (Dent and Flint, 14x7 inbred combinations).

The sRNA population diversity analysis for each of the three hybrids compared to their inbred parents revealed a high variability for the fractions specific to either the hybrid, the inbred lines, or the overlapping fractions respectively (see Figure 15, Figure 16, and Figure 17).





Figure 15: sRNA population diversity of the low heterotic hybrid S028xL024 and its parental inbred lines

Figure 16: sRNA population diversity of the intermediate heterotic hybrid S028xF039 and its parental inbred lines



Figure 17: sRNA population diversity of the high heterotic hybrid P033xF047 and its parental inbred lines

The low heterotic hybrid S028xL024 and the high heterotic hybrid P033xF047 exhibited overall similar ratios of sRNAs specific to either inbred lines or hybrid as well as overlaps of those. The intermediate hybrid S028xF039 differed considerably from the other two. This difference is remarkable as the intermediate and the low heterotic hybrids share the inbred parent S028. 3.2.4.3 Inbred line sRNA population diversity comparison and grouping

Although the first two components of the PCA did only reflect 41.47 % of the variability of the whole data, this information allowed the separation of the inbred lines into their heterotic groups (see Figure 18). The first component of the PCA explained 27.46 % of the variability of the data and primarily separated the inbred lines into their two heterotic groups (Dent and Flint). The second component explained 14.01 % of the variability of the original data and partly separated the heterotic subgroups.





The heterotic groups and subgroups are represented by colors. Flint lines are shown in blue (European Flint in light blue, Flint/Lancaster in dark blue), the Dent lines are shown in red (lodent in light red, lowa Stiff Stalk Synthetic in dark red).

3.2.4.4 Association of inbred parent sRNA expression with hybrid trait-values

The association of differentially expressed sRNAs for the 14x7 inbred parent combinations with high or low heterosis (MPH for GY) in the hybrid offspring resulted in 4,357 positively and 6,915 negatively MPH for GY associated sRNAs, in the following referred to as heterosis-associated sRNAs (ha-sRNAs).

3.2.4.5 Correlation analysis of sRNA-based distance of inbred parents and heterosis The correlation of the (qualitative) binary distance of differentially expressed sRNAs between inbred line combinations resulted in a weak negative correlation with MPH for GY (r=-0.26, see Figure 19). In contrast, the correlation of the (quantitative) euclidean distance, based on expression differences of differentially expressed sRNAs between inbred line combinations, and heterosis in the hybrids resulted in a stronger negative correlation (r=-0.56, see Figure 20).



Figure 19: Correlation of binary distance of differentially expressed sRNAs with heterosis of all 98 hybrids



The correlations of sRNA-based binary distances (D_b) , separately calculated based on pos. and neg. ha-sRNAs respectively for all 14x7 inbred combinations, and MPH for GY of the resulting 98 hybrids, resulted in strong correlations for both subsets. The binary distance D_{b,pos}, calculated from the pos. ha-sRNAs, resulted in stronger correlation with MPH for GY (r=0.85, see Figure 21) than D_{b,neg}, calculated from the neg. ha-sRNAs (r=0.79, see Figure 22). The distribution of D_{b,neg} exhibited two separated groups of in-

r=-0.565 ***

bred combinations, with the majority of inbred combinations grouping with a binary distance below 0.5 and a smaller group with a binary distance above 0.9 (see Figure 22). In contrast to this, the inbred combinations are distributed throughout the distance range of $D_{b,pos}$ (see Figure 21). The correlation of the combined binary distance $D_{b,com}$ from both pos. and neg. ha-sRNAs with MPH for GY reached a stronger correlation (r=0.93, see Figure 23) than the separate correlation of the binary distances calculated from pos. and neg. ha-sRNAs.





Figure 21: Correlation of binary distance of positive ha-sRNAs with heterosis of all 98 hybrids

Figure 22: Correlation of binary distance of negative ha-sRNAs with heterosis of all 98 hybrids



Figure 23: Correlation of binary distance of all ha-sRNAs with heterosis of all 98 hybrids

3.2.4.6 sRNAome-based prediction of hybrid trait values

The prediction accuracy for heterosis was tested for the two hybrid traits heterosis (MPH for GY) as well as hybrid performance (HP for GY) and the two prediction types Type-0 and Type-2 prediction. While the Type-0 prediction does not have any hybrids in the prediction set that have an inbred parent idental to hybrids from the estimation set, in case of Type-2 prediction both inbred parents of each hybrid in the prediction set are mated with different inbred parents in the estimation set.

The prediction accuracy using sRNA-based distance measures was higher and less dispersed for Type-2 hybrids (0.88 \pm 0.1 for MPH for GY, 0.77 \pm 0.09 for HP for GY), than for Type-0 hybrids (0.53 \pm 0.34 for MPH for GY, 0.35 \pm 0.32 for HP for GY) throughout the 100 cross-validation runs (see Figure 24 and Figure 25). The prediction performed better for heterosis (MPH for GY) (see Figure 24) performs better than for hybrid performance (HP for GY), for the latter had a higher fraction of false predictions (see Figure 25).



Figure 24: Prediction accuracy for ha-sRNA predictions of heterosis (MPH for GY)



Figure 25: Prediction accuracy for ha-sRNA based predictions of hybrid performance (HP for GY)

3.2.4.7 ha-sRNA length distribution enrichment analysis

The sRNA length enrichment analysis testing for overrepresentation of specific sRNA lengths for the pos. or neg. ha-sRNAs respectively revealed for both sets a significant enrichment of 22 and 24 nt sRNAs (p<0.001) in comparison to a random distribution (see Figure 26). The pos. and neg. ha-sRNAs differed in the proportion of 22 nt (13.22 % vs. 21 %) and 24 nt sRNAs (64.95 % vs. 56.23 %). All other lengths did not differ between the two sets of ha-sRNAs (see Appendix Table 12).



Figure 26: Length distribution and enrichment analysis of positive and negative ha-sRNAs

The length distribution comparison of mapping position counts for the ha-sRNAs revealed a high number of highly repetitive 22 nt neg. ha-sRNAs with >100 mapping positions on the B73 reference genome (see Figure 27). The number of both pos. and neg. ha-sRNAs that are not mappable to the B73 reference genome is predominant in 24 nt ha-sRNAs and less present in all other sRNA lengths (see Figure 27, Figure 28, Appendix Table 13, and Appendix Table 14).



Figure 27: pos. ha-sRNA length distribution separately for mapping count groups



Figure 28: neg. ha-sRNA length distribution separately for mapping count groups

3.2.4.8 Genome-wide enrichment analysis for ha-sRNAs

The sRNAs as a whole were enriched in the euchromatic arms of all ten chromosomes following both the gene and intergenic and being opposite to repeat genome coverage patterns (see Figure 29).



Figure 29: Genome-wide enrichment analysis for ha-sRNAs

Data shown in the circles: 1) ha-sRNA coverage, 2) sRNA coverage, 3) gene coverage, 4) repeat coverage, 5) intergenic coverage, 6) 22 nt ha-sRNA coverage, 7) enrichment probabilities for 22 nt ha-sRNAs, peaks reaching the green zone show significant enrichment (p<0.05), 8) 24 nt ha-sRNA coverage, 9) enrichment probabilities for 24 nt ha-sRNAs, peaks reaching the green zone show significant enrichment (p<0.05)

The centromeres are highlighted in red in the chromosome axis (outermost circle).

The plot was generated using Circos (version 0.66, Krywinski et al. 2009).

In contrast to this, the ha-sRNAs were evenly distributed throughout the B73 reference genome. A separate consideration of 22 and 24 nt ha-sRNAs revealed that this homogeneous distribution is determined by the 22 nt ha-sRNAs, while the 24 nt ha-sRNAs, in contrast, follow the general trend of depletion around the centromeres and higher fractions of sRNAs in the chromosome arms. This result was supported by the enrichment of ha-sRNAs at centromeric and pericentromeric regions only given for 22 nt neg. ha-sRNAs, thus highly heterochromatic regions which are enriched for repeats.

3.2.4.9 ha-sRNA annotation distribution analysis

Annotating the ha-sRNAs for the annotation types: gene, intergenic, and repeat or all possible overlaps of those types, indicated distinct distribution patterns for ha-sRNA lengths. While ha-sRNAs mapping to single annotation types were mainly 24 nt long, the 22 nt neg. ha-sRNAs constituted the major fractions of the two intersections of annotation types: gene/intergenic and gene/repeat/intergenic (see Figure 30, Appendix Table 15 and Appendix Table 16).



Figure 30: Distribution of ha-sRNAs to annotation types

3.2.4.10 Distribution of ha-sRNAs at and around gene loci

The relative distribution of ha-sRNAs up- or downstream of genes and exon or introns (see Figure 31) indicated a trend towards 24 nt ha-sRNAs constituting the largest fraction of sRNAs close to genes (around 1 kbp up/downstream) as well as close to the splice sites in introns. The 22 nt ha-sRNAs exhibited the opposite pattern with 24 nt ha-sRNA being the largest fraction distant of genes (>2.5 kbp up/downstream) and intron/exon-junctions.



Figure 31: Relative distribution of ha-sRNAs at and adjacent to gene annotations

3.2.4.11 Enrichment/depletion analysis for ha-sRNAs for repeat super-families and families

The enrichment/depletion analysis for ha-sRNAs that are mapping to only one repeat super-family separately for the sRNA lengths 18 nt to 28 nt revealed distinct distributions of 22 nt and 24 nt ha-sRNAs (see Figure 32 and Appendix Table 17) in contrast to a random set of sRNAs (see Figure 33, Appendix Table 18). The 22 nt ha-sRNAs exhibited a significant enrichment of LTR Gypsy retrotransposons (p<0.001), while the 24 nt ha-sRNAs were significantly enriched for DNA transposons of the super-families TIR CACTA, TIR hAT, TIR Mutator and TIR Pif/Harbinger (p<0.001) as well as the LTR Unknown retrotransposons (p<0.05). The LTR Unknown retrotransposon was significantly enriched for 22 nt ha-sRNAs (p<0.001). The LTR Copia retrotransposons were significantly depleted for 22 nt ha-sRNAs (p<0.05) (see Figure 34, Appendix Table 19).



Figure 32: sRNA length distribution of ha-sRNAs to repeat super-families



Figure 33: sRNA length distribution of bootstrap sample average to repeat superfamilies

	18	19	20	21	22	23	24	25	26	27	28
LINE L1											
LINE Unknown					***	▼					
LTR Copia	***	***	***	••	***	••	***	••	••	••	▼
LTR Gypsy	***	***	***				***	••	***	***	••
LTR Unknown					***						
TIR CACTA	***	***	***	***	***			••	***	••	▼
TIR hAT	••	***	***	***	***	▼		••	••		
TIR Mutator			▼	▼	***				▼		
TIR Pif/Harbinger	•	••	••	***	***						
TIR Tc1/Mariner					***						

Figure 34: ha-sRNA enrichment and depletion probabilities

significance of enrichment: $\blacklozenge p <= 0.05 \land \blacktriangle p <= 0.01 \land \blacktriangle p <= 0.001$ significance of depletion: $\lor p <= 0.05 \lor \lor p <= 0.01 \lor \lor \lor p <= 0.001$

The enrichment/depletion analysis performed separately for pos. and neg. ha-sRNAs lengths revealed a higher fraction of 22 nt neg. ha-sRNAs than pos. ha-sRNAs (see Figure 35 and Figure 36, Appendix Table 20 and Appendix Table 21). The 22 nt neg. ha-sRNAs were significantly depleted for all repeat super-families except for LTR Gypsy retrotransposons (see Figure 34). Most enrichments for ha-sRNA lengths were identical between pos. and neg. ha-sRNAs, except for the enrichment of 21 nt neg. ha-sRNA for LTR retrotransposons, 24 nt pos. ha-sRNAs for Pif/Harbinger DNA transposons, and 21 and 24 nt pos. ha-sRNAs for LTR Unknown retrotransposons not present in the ha-

sRNAs with opposed association (see Figure 37 and Figure 38, Appendix Table 22 and Appendix Table 23).



Figure 35: sRNA length distribution of pos. ha-sRNAs to repeat super-families



Figure 36: sRNA length distribution of neg. ha-sRNAs to repeat super-families



	18	19	20	21	22	23	24	25	26	27	28
LINE L1											
LINE Unknown					••						
LTR Copia	***	***	***	***	***	***	***	***	••	***	••
LTR Gypsy	***	***	▼				***	***	***	***	***
LTR Unknown					***						
TIR CACTA	••	***	***	***	***			••	••	•	
TIR hAT	••		••	***	***				▼		
TIR Mutator				••	***						
TIR Pif/Harbinger		••	••	▼	***						
TIR Tc1/Mariner					••		V				

Figure 38: neg. ha-sRNA enrichment and depletion probabilities to repeat super-families

significance of enrichment: $\blacklozenge p <= 0.05 \land \blacktriangle p <= 0.01 \land \blacktriangle p <= 0.001$ significance of depletion: $\lor p <= 0.05 \lor \lor p <= 0.01 \lor \lor \lor p <= 0.001$

Figure 37: pos. ha-sRNA enrichment and depletion probabilities to repeat super-families

The enrichment/depletion analysis of ha-sRNAs to repeat families, with at least 50 hasRNAs mapping to the repeat family (see Figure 39, Appendix Table 24), revealed enrichments/depletions for various retrotransposon families (see Figure 40, Appendix Table 25). In contrast to the repeat super-families with enrichments for 24 nt only given for DNA transposons, their enrichments for the LTR Unknown families ipiki and yraj were found. While 22 nt ha-sRNAs were only enriched for the LTR Gypsy repeat superfamily, their enrichment in repeat families was found for the LTR Copia family leviathan besides various LTR Gypsy families, namely cinful-zeon, gyma, and nihep. Contrasting to the repeat super-family results, no enrichments were found for all ha-sRNAs to DNA transposons.



Figure 39: sRNA length distribution of ha-sRNAs to highly expressed repeat families

	18	19	20	21	22	23	24	25	26	27	28
RLC ji	•	•	***	•			***	▼		▼	
RLC leviathan											
RLC opie		••	••				***	••		•	
RLG cinful-zeon		•					••				
RLG doke							••				
RLG gyma							•				•
RLG huck	••	••	••	▼			••	••		V	
RLG nihep						•					
RLG prem1											
RLG xilon-diguus							***				
RLX gunu		•					***		•		
RLX ipiki											
RLX milt				••	***						
RLX yraj					***						

Figure 40: ha-sRNA enrichment and depletion probabilities to repeat families



The enrichment of sRNA lengths of pos. ha-sRNAs for repeat families (see Figure 41, and Figure 42, Appendix Table 26, and Appendix Table 27) revealed 22 nt to be enriched for the LTR Copia family ji. The DNA transposon TIR Mutator family ZM00118 exhibited an enrichment for 24 nt pos. ha-sRNAs and in contrast the family Zm00754 for 23 nt pos. ha-sRNAs.



Figure 41: sRNA length distribution of pos. ha-sRNAs to highly expressed repeat families

18 19 20 21 22 23 24 25 26 27 28

Figure 42: pos. ha-sRNA enrichment and depletion probabilities to repeat families

RLC ji						▼			
RLG cinful-zeon						***			
RLX ipiki									
TIR Mutator Zm00118									
TIR Mutator Zm00754									
significance of enrichr significance of depleti	nent on:	a ▲ : • • •)<=()<=(0,05 0,05	p<= p<=	0,01	▲ p< ▼ p<	=0,0)01)01

The neg. ha-sRNAs were found to feature the most diverse enrichment patterns. Various retrotransposon families showed enrichments for multiple sRNA lengths (see Figure 43 and Figure 44, Appendix Table 28 and Appendix Table 29). Consistent with all hasRNAs, the 22 nt neg. ha-sRNAs showed an enrichment for the LTR Copia retrotransposon family leviathan as well as the LTR Gypsy families cinful-zeon, gyma, nihep, and xilon-diguus.



Figure 43: sRNA length distribution of neg. has sRNAs to highly expressed repeat families

	18	19	20	21	22	23	24	25	26	27	28
RLC ji	▼	▼	••	••	***		••				
RLC leviathan											
RLC opie		▼	▼		***	▼	••	▼			
RLG cinful-zeon											
RLG gyma											
RLG huck		▼	••		***		***	▼	▼		
RLG nihep											
RLG xilon-diguus							▼				
RLX ipiki											
RLX yraj					▼						
significance of enrichment: $\blacktriangle p <= 0.05 \land \blacktriangle p <= 0.01 \land \blacktriangle p <= 0.001$ significance of depletion: $\forall p <= 0.05 \forall \forall p <= 0.01 \forall \forall \forall p <= 0.001$											

Figure 44: neg. ha-sRNA enrichment and depletion probabilities to repeat families

The enrichment of ha-sRNAs for the LTR Unknown ipiki retrotransposon family for all sRNA lengths was found to be in conformity of nearly all sRNA lengths of the LTR Unknown super-family (see Figure 34 and Figure 44 in comparison).

3.2.4.12 ha-sRNA expression pattern analysis

The expression patterns of differentially expressed sRNAs as well as ha-sRNAs were tested for changes in the hybrids by d/a expression pattern analysis. The analysis revealed a bias of sRNA expression towards low-parent-like (d/a=-1) and mid-parent or additive (d/a=0) expression in the hybrids. Expression extremes (below low parent d/a<1 and above high parent d/a>1) were less represented (see Figure 45). This ex-

pression pattern was consistent for all differentially expressed sRNAs, as well as pos. ha-sRNAs and neg. ha-sRNAs separately.



Figure 45: Dominance/additivity (d/a) expression pattern analysis of differentially expressed sRNAs and ha-sRNAs

The comparison of 22 and 24 nt ha-sRNAs, separately for neg. and pos. associated subsets, for the ratio of fold-change to presence-absence expression patterns in the 14x7 inbred combinations revealed a major trend for presence-absence expression pat-

terns towards the neg. ha-sRNAs in comparison to the pos. ha-sRNAs. The pos. ha-sRNAs had a trend towards equal amounts of fold-change and presence-absence expression patterns (see Figure 46).

Figure 46: Expression pattern ratio for fold-change and presence-absence variation of 22 nt and 24 nt ha-sRNAs based on 14x7 inbred combinations



3.2.5 Analysis of effects on the transcriptome mediated by ha-sRNAs

3.2.5.1 ha-sRNA target prediction

The target prediction resulted for 8,626 of the 11,272 ha-sRNAs in putative targets that are potentially regulated by post-transcriptional gene silencing or inhibition of translation. The number of ha-sRNA/target-relations was revealed to be proportional to the number of ha-sRNAs for each sRNA length group. The mean ratio of targets per ha-sRNA was found to differ between the sRNA lengths. The 18 nt and 19 nt ha-sRNAs were shown to exhibit a very high ratio of ~10 and ~7 targets per ha-sRNAs, while the ratio for other sRNA lengths varied between 1.7 to ~4.5 targets per ha-sRNA (see Table 3).

sRNA length [nt]	number of ha-sRNAs with target	total number of distinct targets	targets per ha-sRNA
18	137	1346	9,82
19	133	897	6.74
20	122	484	3,97
21	477	2,140	4,49
22	1,642	6,398	3,9
23	512	1,783	3,48
24	4,710	12,664	2,69
25	170	453	2,66
26	104	226	2,17
27	109	188	1,72
28	122	235	1,92
all (18-28)	7,965	27,879	3,5

Table 3: Result summar	y of ha-sRNA targe	t prediction with	psRNATarget

3.2.5.2 Correlation of ha-sRNA with microarray expression data

The correlation of ha-sRNA expression with target transcriptome expression data generated on the 46k maizearray was performed for 206,565 ha-sRNA ha-sRNA/target combinations to reveal mRNA transcripts putatively linear regulated by ha-sRNAs. After correction for multiple testing, 22 transcripts were identified to show a significant correlation with 38 ha-sRNAs. The separate consideration of pos. and neg. ha-sRNAs revealed 15 target transcripts significantly correlated with 22 pos. ha-sRNAs and four transcripts with significantly correlation with 18 neg. ha-sRNAs.
The length distribution of ha-sRNAs with correlation to putative target transcripts was found to exhibit a clear trend towards 21 nt and 22 nt ha-sRNAs (see Figure 47, Appendix Table 30).

30 all ha-sRNAs 25 pos. ha-sRNAs fraction of sRNAs [%] neg. ha-sRNAs 20 വ 9 വ Figure 47: Length distribution of ha-sRNAs 0 with expression significantly correlated to puta-18 19 20 21 22 23 24 25 26 27 28 tive target transcripts sRNA length [nt]

3.2.5.3 sRNA differential expression analysis

In total, 4,808 of the ha-sRNAs were present in the sRNA dataset by Regulski *et al.* (2013). 109 of those sRNAs were found to be differentially expressed between B73 and Mo17. The size distribution of the differentially expressed ha-sRNAs was found to exhibit a lower fraction of 24 nt ha-sRNAs but higher fractions of non-specific sRNAs with length >25 nt than all ha-sRNAs. The relative fraction of 20 nt to 22 nt ha-sRNAs was shown to be nearly identical for all as well as the differentially expressed ha-sRNAs (see Figure 48, Appendix Table 31).





3.2.5.4 RNAseq differential expression analysis

The RNAseq reads were assembled to transcripts for a following test for ha-sRNAs present in the identical biological sample that putatively regulate these transcripts. The transcriptome assembly of the two inbred lines B73 and Mo17 resulted in 114,082 genes. A subset of 7,008 was found to be differentially expressed.

3.2.5.5 Enrichment analysis for ha-sRNAs at differentially expressed transcript genome regions

The enrichment analysis for ha-sRNAs upstream, within, and downstream of differentially expressed genes determined from the RNAseq transcriptome assembly resulted in differing enrichments for 22 and 24 nt ha-sRNAs. While 22 nt ha-sRNAs were predominantly located in the gene annotation and less in 1kbp regions up- and downstream, the 24 nt ha-sRNAs were evenly distributed throughout these regions (see Table 4).

Table 4: Distribution of ha-sRNAs of specific length to regions of differentially expressed genes

gene region	18 – 28 nt	22 nt	24 nt
upstream	71	13	57
gene	67	55	53
downstream	65	16	44

4. Discussion

4.1 Re-annotation of the 46k-maizearray

Since their introduction in 1995, DNA microarrays have become a mature gene expression analysis technology for the analysis of gene sets, which were assembled based on prior selection (Sheena et al. 1995). The microarray technology is based on DNA probes (oligonucleotides) with sequences complementary to genes of interest, which are synthesized or cross-linked to a solid surface. The oligonucleotides are hybridized with fluorescent-labeled cDNAs or RNAs, which are generated from transcript samples (Phimister 1999). A few microarray platforms have been published for genome-wide expression analysis in maize, including the 57k and 46k maize oligonucleotide arrays based on long-oligonucleotides (~70 nt) from the Maize Oligonucleotide Array Project (Gardiner et al. 2005). These two microarray platforms have served for the expression analysis in various studies until these days (see Appendix Table 1), although gene expression analyses in most recent studies are performed by RNA high throughput sequencing. The analysis of microarray experiments is highly reliant on the quality of the design of the oligonucleotide probes and the integrity of the annotation. The latter did, in case of the maize oligonucleotide arrays, not include information from the B73 reference genome and thus required a re-annotation to improve gene expression analyses and allow for the identification of putative target effects of sRNAs.

The re-annotation of the 46k-maizearray resulted primarily in a qualitatively improved annotation of the oligonucleotide probes. The reason for the improvement of the re-annotation in contrast to the official annotation (version 4; www.maizearray.org), which was mainly based on ESTs (Gardiner *et al.* 2005), can be mainly objected to the availability of B73 reference genome sequence and annotation. The decrease of the number of annotated probes is caused by the identification of probes that were not optimally localized to the intended genes of interest during the design of the microarray and thus might result in misinterpretation of analyses depending on the official annotation. The mispositioned oligonucleotides antisense to protein coding genes are most probably caused by misoriented ESTs or cover natural antisense transcripts (NATs) (Jin *et al.* 2008). Other oligonucleotides positioned antisense to transposable elements putatively reflect dsRNA intermediates of RNAi pathways (Ito 2012). In contrast, oligonucleotides representing multiple genomic loci, thus more than one gene locus, reflect copies of

genes generated from genome duplication events or transposition (Schnable *et al.* 2009). Although some oligonucleotides were not located in the optimal position, more than half of the oligonucleotides represent single genes. The re-annotation especially improved the functional annotation, resulting in nearly twice as many probe targets with GO-terms. The information about unintentional cross-hybridizations, identified by the mapping of the oligonucleotides to the maize reference genome, and the enhanced functional annotation will enable a more precise analysis of gene expression profiles generated on this array platform. The re-annotation procedure is transferable to other microarray platforms of organisms with improved genome sequence assemblies and annotation since the design of the microarray probes.

4.2 sRNA sequencing data normalization, the neglected factor

High throughput sequencing technologies have evolved into key technologies for expression analysis. This is founded, among other factors, such as labor-insensitivity and cost, in the independence of prior knowledge about the sequences of interest, sequence information with nucleotide-resolution, and a dynamic quantitative signal range without saturation or detection limits. The characterization of plant sRNA transcriptomes is nowadays based on sRNA sequencing as shown by a multitude of studies (McCormick *et al.* 2011).

When comparing sRNA expression between different conditions or lines, it is critical to perform an appropriate normalization of the sRNA expression data due to differing sequencing depths which are depending on the sequencing technology (McCormick *et al.* 2011, Dillies *et al.* 2012). Although normalization is important for the outcome of expression analyses, particularly the data normalization in sRNA transcriptome studies was often neglected. In most cases, it was confined to scaling the sequence counts by a fixed factor defined by the sequencing depth (Bolstad *et al.* 2003). This practice has a major shortcoming which is substantial in the case of differing sequencing depths. As an example, we consider two samples A and B of different sequencing depths and assume that $n_A=4,000,000$ and $n_B=6,000,000$ are the number of total reads obtained from an sRNA sequencing library would assume the samples. A and B to be equal, regarding their composition, with the only difference that each sRNA was sequenced 1.5-fold in sample B relative to sample A. Ideally, the sequencing of sRNAs is fully independent of the

sample preparation steps and the sequencing and thus equal for both samples (McCormick et al. 2011). The problem in the comparison of differing sequencing depths is that sequencing experiments do represent a stochastic subset of the sample and not an exhaustive result. This implies that the sequenced subset of sRNAs does not reflect the actual set of sRNAs in the sample. Due to the stochastic sampling, the high abundant sRNAs are likely to be present in the sequencing sample, but not the lowabundant. Thus, relatively more low-abundant sRNAs will be sequenced at higher sequencing depths. In the aforementioned example, sample B will contain low-abundant sequences not covered by the lower sequencing complexity of sample A (Robinson & Oshlack 2010). Thus, the scaling factor of B to A is not expected to be 1.5 but smaller. In addition, especially the expression values of low abundant samples will most likely not be in a linear relation to the sequencing depth. The slightly modified quantile normalization, which was applied to the sRNA expression data from the inbred lines and hybrids of the breeding program of the University of Hohenheim and the reference inbred line B73, assumes an overall predominant non-differential expression between the lines and results in an adjustment of the expression values resulting in an equal distribution for all lines (Bolstad et al. 2003). It should be noted that, due to the assumption of primarily non-differential expression, this normalization method should result in an underestimation of differential expression and is thus expected to result in a larger number of false negatives. But more important, it should not introduce false positives, as a scaling by sequencing depth unintentionally does.

The problem of sequencing data normalization gets even more complex if any differing conditions between the two samples results in different compositions of sequences. This was unknown in the comparison of the induced/non-induced transgenic P1/HC-Pro hybrid plants, as effects of the gene silencing suppressor could putatively affect single sRNA biogenesis pathways and thus the normalization would introduce a bias into the final expression data. For these samples, the expression was calculated through bootstrapping with replacement of fixed subsets from the original sample. This normalization relies solely on the sequencing data and does not imply any assumptions about the expression of the sRNAs.

The normalization of sequencing data of inbred lines of large breeding populations is expected to be more efficiently tractable and be less demanding regarding to computational time and memory consumption if the samples are individually normalized via the bootstrap-based normalization method. This would allow the introduction of additional lines into analyses, without the re-normalization of the whole set of lines of the breeding program, as it would be mandatory for the quantile normalization.

4.3 Functional identification of sRNA effects on heterosis formation using a viral gene silencing suppressor

In this work, sequencing of sRNA populations of induced and non-induced transgenic maize hybrids expressing the P1/HC-Pro gene silencing suppressor was performed to analyze the suppressor effects on the sRNA populations of these plants. It was shown that hybrid maize plants expressing the P1/HC-Pro gene silencing suppressor exhibit an increase in heterosis for growth rate (Thiemann 2011). The induction of P1/HC-Pro expression was performed early in development (1 dap), thus early effects regulated by sRNAs due to interplay of the two parental genomes, e.g. on DNA methylation patterns, are affected by the gene silencing suppressor.

Although the gene silencing suppressor P1/HC-Pro from *WSMV* has been shown to interfere with RNA silencing and P1, rather than HC-Pro is known to be the active protein component, the exact functional mechanism is unknown (Young *et al.* 2012). This activity is in contrast to the virus family Potyviridae with HC-Pro mediating suppressor activity by binding 21 nt sRNAs (Young *et al.* 2012). Thus, the analysis of sRNAs populations of transgenic plants containing the WSMV P1/HC-Pro was furthermore performed to elucidate this mechanism and explain how the heterosis in these transgenic plants is established.

A general effect of the P1/HC-Pro gene silencing suppressor resulting in a reduction by ~30 % of the sRNA population complexity in induced transgenic hybrids could be revealed by bootstrap analysis. The reduction of the sRNA population complexity suggests that the suppressor either acts in the biogenesis of sRNAs or leads to a faster degradation of sRNAs after their generation, by either directly promoting degradation of the sRNAs or preventing their stabilization. Interactions of P1/HC-Pro with effector complexes (e.g. AGO) can be excluded based on this result, as changes in the sRNA population complexity would not be constituted by effector proteins.

The suppressor activity does not result in a significant change of the sRNA length distribution in induced vs. non-induced transgenic plants. The decrease of sRNA fractions of 22 nt and 24 nt is too marginal for the assumption that P1/HC-Pro affects specifically the biogenesis of these sRNAs. Thus P1/HC-Pro is not expected to interfere with components in specific biogenesis pathways, e.g. DCL proteins, which determine the generation of well-defined sRNA lengths. This is consistent with the result that miRNAs were shown not to be differentially expressed between induced and non-induced hybrids.

In combination, the reduced sRNA population complexity without any observable constriction to a specific sRNA length indicates an overall non-specific reduction of sRNAs. This non-specific reduction is expected to evolve either from the interference of P1/HC-Pro with components involved in multiple sRNA biogenesis pathways or sRNA stabilization or by directly acting on the sRNAs by unspecific degradation or competitive binding. Any of these interference mechanisms would be expected to result in the reduction of expression levels for all sRNAs. Thereby, the removal of sRNAs from the sRNA population in the induced transgenic P1/HC-Pro plants, which are low abundant in the non-transgenic P1/HC-Pro plants, leads to a lower sRNA population complexity.

The increased heterosis for growth rate for induced transgenic P1/HC-Pro plants (Thiemann 2011) exhibiting a less complex sRNA population in contrast to non-induced plants reveals sRNAs as a factor in the formation of heterosis, instead of solely being a consequence of heterosis-related mechanisms. A previous study analyzing maize plants deficient for Mediator of Paramutation 1 (MOP1), a RDR2 orthologue, resulting in the disruption of 24 nt siRNA biogenesis, revealed negative effects on various traits (days to 50 % anthesis, cob weight, stover biomass) in inbred lines, but an increase in hybrid performance and heterosis of these developmental and yield related traits relative to the wild-type plants (Barber et al. 2012). The authors thus conclude that 24 nt siRNAs dependent on MOP1 are not required for heterosis formation of the measured traits in the tested lines (Barber et al. 2012). Another study analyzed the Arabidopsis HEN1 mutant, which does not methylate sRNAs which results in the degradation of those unprotected sRNAs (Shen et al. 2012). Both HEN1 mutant inbreds and hybrids show a dwarfed phenotype suggesting a severe developmental defect by perturbing miRNAs in plant development (Groszmann et al. 2013). Shen et al. (2012) hypothesize that the reduced hybrid performance and heterosis are an effect of lost DNA methylation due to the lack of sRNAs. The study does not provide information about the effects on the sRNA population complexity by HEN1, but on the reduction of sRNA cluster expression in this mutant (Shen *et al.* 2012). Groszmann *et al.* (2013) assume that this severe developmental phenotype masks the effect of heterosis and thus does not allow any conclusion of the effect of sRNAs on heterosis. Although the reduction of miRNAs might affect the development and thus the hybrid performance is expected to be lower than in wild-type plants, heterosis is not assumed to be dependent of developmental processes or the regulation by miRNAs (East 1936, Barber *et al.* 2012). The P1/HC-Pro plants do not exhibit an altered phenotype in the analyzed seedling stage. This suggests that in comparison to the HEN1 maize mutant, P1/HC-Pro less strictly interferes with the affected RNA silencing components or directly with sRNAs.

4.4 sRNA transcriptome analysis in inbred lines from a breeding population and hybrids of different heterosis-levels

The sRNA expression differences between inbred lines as well as their hybrids discovered for various plant species, as discussed in the previous chapter, suggest sRNAs to be regulators involved in heterosis formation. This hypothesis is supported by the effect on heterosis observed in the MOP1 mutant (Barber *et al.* 2012) and P1/HC-Pro transgenic plants.

4.4.1 sRNA transcriptomes reflect the genetic diversity of maize inbred lines

Maize exhibits extraordinary levels of genetic diversity with differences between inbred lines similar to differences between human and chimpanzee (Springer *et al.* 2009). Maize inbred lines as well as hybrids exhibit a large amount of sequence variation as SNP, short indels, PAV, CNV to a different extent and a considerable non-collinearity in the genomes of different maize inbred lines (Tenaillon *et al.* 2001, Song & Messing 2003, Hochholdinger & Hoecker 2007, Springer *et al.* 2009, Hansey *et al.* 2012). Another source of sequence variation in maize is contributed by transposable elements, especially retrotransposons, which differ in their dispersion and copy number between haplotypes (Baucom *et al.* 2009).

All previous studies analyzing sRNA expression between inbred lines and their hybrids, in various plant species, analyzed only a single inbred line pair and their hybrids (Ha *et al.* 2009, Chen *et al.* 2010, He *et al.* 2010, Groszmann *et al.* 2011, Barber *et al.* 2012,

Chodavarapu *et al.* 2012, Li *et al.* 2012, Shen *et al.* 2012, He *et al.* 2013). It is currently unknown if there are specific sRNAs species or pathways that have effects in heterosis formation. In this work, sRNA transcriptomes of inbred lines from a maize breeding program were analyzed for their diversity and by association to heterosis of their hybrids to elucidate their contribution to the formation of heterosis.

The genetic diversity of elite maize inbred lines could be reflected by sRNA transcriptomes in terms of fractions of sRNA reads that are mappable to the B73 reference genome. With increasing diversity the number of mappable sRNAs decreases due to the extent of differences to the reference genome. As expected the sRNA transcriptome of B73 exhibits the largest fraction of mappable sequences to the B73 reference genome. It is noteworthy that not all sRNAs from the B73 sample could be mapped to the reference genome generated from this line. The fraction of unmappable reads might either reflect sequencing errors or regions of the genome that are not properly assembled due to low sequence coverage or high complexity as given for centromeric regions containing repetitive elements (Zhou et al. 2009, Barthelson et al. 2011, Schatz et al. 2012). The sRNA populations of all other 21 inbred lines have fractions of about 25 % to 40 % of reads that could not be localized on the reference genome. The sRNA transcriptomes of the three analyzed hybrids hold significantly higher fractions of sequences located on the reference genome in comparison to the inbred lines. Thus, the expression of these sRNAs that are differential in their presence between inbred lines and the B73 reference might be lost in hybrids.

Sequence variations exhibit different levels of conservation that can be characterized through the concept of the core genome and the dispensable genome that together constitute the pan-genome (Medini *et al.* 2005, Gore *et al.* 2009, Hirsch *et al.* 2014). While the core genome reflects the conserved part that is present in a large number of lines, the dispensable genome describes the variable part. The analysis of the sRNA population diversity between pairs of inbred lines of 14x7 resulting combinations shows that the intersection of the sRNA pan-transcriptomes of pairs of inbred lines only covers about a third of the sRNAs of both lines. Thus, sRNA populations of maize inbred lines of two heterotic groups feature a high intraspecific variability. Additional variability for sRNA populations was revealed in three inbred-hybrid triplets between the inbred parents and their hybrid offspring. The fraction of sRNAs specific to the hybrid was smaller in the study of Barber *et al.* (2012) in comparison to the results of this study. The differ-

ence can be attributed to differences in the analyses, as Barber *et al.* (2012) only considered mapped sequences and one hybrid. The consideration of core and dispensable sRNA transcriptomes is supported by the fractions of mapping sRNAs to the B73 reference genome. These results show that the dispensable sRNA transcriptome is larger in inbred lines than in hybrids that exhibited low fractions of unmapped reads similar to the B73 reference sample.

Heterosis results from heterozygosity of the combination of two divergent genomes and thus the combination of variation at multiple genomic loci (Springer & Stupar 2007). It was suggested that differential gene expression between parental lines and hybrids results in heterosis (Song & Messing 2003, Guo *et al.* 2006, Stupar & Springer 2006, Swanson-Wagner *et al.* 2006, Meyer *et al.* 2007, Hoecker *et al.* 2008, Thiemann *et al.* 2010). This differential gene expression might originate from epigenetic regulations like siRNAs and novel or altered epigenetic states generated from epigenetically diverse inbred parents.

This work shows that sRNAs are able to reflect the genetic distance of inbred lines, not only within but moreover between genetically divergent heterotic groups by the successful separation of inbred lines into their heterotic groups based on differentially expressed sRNAs between inbred lines. The heterotic groups represent two genetically separated groups of germplasm, and thus exhibit a higher genetic distance between than within the heterotic groups (Reif et al. 2005). Previous studies have shown that genetic distance estimates of inbreds based on molecular markers are closely related to heterosis for crosses of related lines or lines of the identical heterotic group, but not for crosses between lines of genetically divergent heterotic groups (Melchinger 1999, Cheres et al. 2000). Nevertheless, differentially expressed parental sRNAs that discriminate the inbred lines in their heterotic groups and thus reflecting the variability of those lines, should be related to heterosis in the F1 offspring. The ability of inbred line grouping suggests that the differentially expressed sRNAs from the combination of the dispensable sRNA transcriptome, exhibiting presence-absence pattern between inbred parents, and a subset of the core sRNA transcriptome, qualified by differential expression between the inbred parents, are able to reflect the genetic distance of these lines. It should be noted that the generation of sRNAs from genomic loci depends on their epigenetic constitution in the inbred lines that defines distinct epialleles between the different lines (Eichten et al. 2013). Thus, sRNAs provide extended information about the epigenetic

status (e.g. DNA-methylation) of the inbred lines, which results in different expression patterns of sRNAs. The importance of knowledge about epigenetic differences between different lines was highlighted by analyses of two *Arabidopsis thaliana* ecotypes that do only differ in their DNA-methylation patterns and only to a very low degree in their genome sequence, but still exhibit high heterosis (Fujimoto *et al.* 2011, Shen *et al.* 2012). The information about epialleles and their dynamic is not reflected by any genetic markers relying solely on the genome sequence. Although it is possible to obtain information about the DNA methylation state by sequencing of bisulfite-treated DNA (Frommer *et al.* 1992), this would not provide information about the parental sRNA expression levels that result in an alteration of the DNA methylation patterns after hybridization. Furthermore, the analysis of epigenetic states by DNA methylation sequencing in contrast to sRNA sequencing data depends on the existence of a reference genome and does not provide information about the dynamics of epialleles.

4.4.2 ha-sRNAs act antagonistically in heterosis formation

The association of sRNA differential expression between inbred parents with heterosis in their resulting hybrids was performed on a set of inbred lines and hybrid field data, which was previously analyzed for the microarray-based identification of heterosis correlated genes (Thiemann et al. 2010). This allows both the comparison of the different levels of regulation of heterosis and the identification of the effects of ha-sRNAs to the mRNA transcriptome. The analysis revealed a larger fraction of neg. ha-sRNAs. Although the association does not provide information about the quantitative effect of the distinct ha-sRNAs on heterosis (MPH for GY), the fact that the majority of ha-sRNAs is related to low heterosis suggests a restraining effect of sRNAs on heterosis. This hypothesis is supported by the increase in heterosis due to the reduction of the sRNA population complexity as shown in the P1/HC-Pro experiment (see chapter 3.2.3). The mainly restraining effect of sRNAs was furthermore revealed by correlation of sRNA distance measures, either qualitative (binary distance) or quantitative (euclidean distance) for all differentially expressed sRNAs with heterosis (MPH for GY). While the qualitative distance considers the number of differentially expressed sRNAs between the inbred parents, the quantitative distance is based on the expression differences of the differentially expressed sRNAs of the parental lines. Both correlations result in a negative relation of differential sRNA expression and heterosis, confirming a predominant restraining

effect. Correlation analyses based on binary distance separately for the pos. or neg. hasRNAs result in considerably stronger correlations than the correlations based on binary distance of unspecific differentially expressed sRNAs without the identification of a statistically significant association to heterosis, suggesting a highly specific effect of the hasRNAs on heterosis. This finding is furthermore supported by a highly accurate sRNAbased prediction of heterosis. The combination of pos. and neg. ha-sRNAs results in the highest correlation with MPH for GY, suggesting that the ha-sRNAs are acting antagonistically in heterosis formation.

These findings in combination with the gene silencing suppressor experiment lead to the hypothesis of mainly negative regulatory effects to be conducted by ha-sRNAs. The gene silencing suppressor experiment resulted in a reduction of the sRNA population complexity of ~30 % which is assumed to eliminate a larger fraction of neg. ha-sRNAs, and thus results in the observed ~15 % increase of growth heterosis.

4.4.3 ha-sRNA characterization reveals enrichment of specific sRNA features

Interestingly, the analysis of ha-sRNAs for known maize miRNAs (miRBase release 20; Griffiths-Jones *et al.* 2006) did not reveal miRNAs to be involved in heterosis formation. Previous studies did not reveal consistent differential expression patterns between inbred lines but between parents and their resulting hybrids (Barber *et al.* 2012, Ding *et al.* 2012, Li *et al.* 2014). Differing expression patterns were shown to solely reflect differences in the rate of development (Barber *et al.* 2012), while no dramatic alterations to the developmental program of hybrids exist (Birchler *et al.* 2010). This result is consistent with East (1936), who hypothesized that developmental genes should not be important for heterosis.

The length distribution of ha-sRNAs uncovers a primarily composition of 22 and 24 nt sRNAs. A significant enrichment of these two sRNA lengths could be shown for the ha-sRNAs, suggesting a specific generation of the ha-sRNAs and not solely reflecting a subset from the sRNA populations. While the enrichment of pos. ha-sRNAs is higher for 24 nt sRNAs, the neg. ha-sRNAs exhibit a higher amount of 22 nt sRNAs. Both the 22 and 24 nt siRNAs are known to be involved in RNA-dependent DNA methylation (RdDM) pathways via the RDR6-RdDM and Pol IV-RdDM respectively (Nuthikattu *et al.* 2013, Matzke & Mosher 2014). The model of the Pol IV-RdDM pathway starts with the

generation of aberrant RNA transcripts, e.g. lacking 5' cap or 3' poly(A) tail, by the plantspecific DNA-dependent RNA polymerase IV (Mosher et al. 2008). These aberrant transcripts are transformed into dsRNA by RDR2 and subsequently cleaved into 24 nt siRNAs by DCL3 (Gazzani et al. 2004, Daxinger et al 2009). The siRNAs are incorporated into RISCs containing an AGO4 protein which directs this complex to nascent Pol V transcripts, where the complex guides other protein components to effect de novo DNA or histone methylation (Wierzbicki et al. 2008). More recently, it was shown that the RDR6-RdDM pathway depends on transcripts generated from Pol II (Nuthikattu et al. 2013). These transcripts, mainly generated from young TEs, are turned into dsRNA by RDR6 (Matzke & Mosher 2014). The dsRNA is processed by either DCL2 or DCL4 resulting in 21 nt to 22 nt siRNAs which are directing DNA methylation at loci with Pol V scaffold transcripts, guided by AGO2 containing RISCs (Wu et al. 2012, Matzke & Mosher 2014). The RDR6-RdDM mechanism was proposed to act as an initiating process for silencing of transcriptionally active TEs, subsequently directing reinforcement of TE methylation via the Pol IV-RdDM pathway (Nuthikattu et al. 2013). The relevance of the RDR6-RdDM mechanism for the initiation of TE silencing might explain the enrichment of 22 nt ha-sRNAs as essential for the maintenance of epialleles with favorable effect on heterosis.

The analysis of mapping position counts of the ha-sRNAs to the B73 reference genome, separately for the pos. and neg. ha-sRNAs, reveals an enrichment of highly repetitive (>100 mapping positions) 22 nt neg. ha-sRNAs. The comparison of all other mapping position count groups did not reveal any other drastic differences between the pos. and neg. ha-sRNAs, thus there is no obvious difference between pos. and neg. ha-sRNAs based on different copy numbers of their generation loci. Interestingly, the non-conserved sRNAs not mapping to the B73 reference genome state the largest amount of ha-sRNAs, highlighting the importance of genetic diversity between inbred lines for the establishment of heterosis. In contrast, ha-sRNAs mapping uniquely or up to ten times to the reference genome make up a large fraction of the 24 nt ha-sRNAs. The mapping position count groups with more than 11 mapping positions on the reference genome, containing more 22 nt ha-sRNAs than 24 nt ha-sRNAs, exhibit an opposite composition in comparison to the groups covering sRNAs with lower copy numbers. This suggests differences in the biogenesis and targeting for the 22 nt and 24 nt ha-sRNAs.

while in contrast to this, the 24 nt ha-sRNAs exhibit a higher specificity. Thus, 22 nt hasRNAs might have a genome-wide effect on epialleles, stochastically affecting gene expression through spreading of DNA methylation at some of the affected loci (Eichten *et al.* 2012). This mechanism fits with the concept, suggested by McClintock (1984), of transposable elements acting as controlling elements not only by transposition but also through imposing its repressive epigenetic state on neighboring genomic regions (Hollister & Gaut 2009).

The distribution of ha-sRNAs exhibits an enrichment of neg. 22 nt ha-sRNAs at pericentromeric regions of all 10 maize chromosomes, while all other lengths of the hasRNAs do not show specific enrichment patterns to genomic regions. Analyses of the maize nested association mapping (NAM) population revealed that the pericentromeric regions contain a 30 % higher residual heterozygosity than regions outside of centromeres, while covering 21 % of the total genic fraction (McMullen et al. 2009, Gore et al. 2009). The Hill-Robertson effect postulates a selective pressure to maintain heterozygosity in regions with low recombination as pericentromeric regions. This effect is caused by the repulsion phase linkage of highly favorable dominant alleles with detrimental alleles (Hill & Robertson 1966, McMullen et al. 2009). An extended consideration of the Hill-Robertson effect not only limited to linkage of favorable to detrimental genes, but also to detrimental regulatory elements, suggests the neg. 22 nt ha-sRNAs to be these detrimental regulatory elements, which result in pseudo-overdominance for the dominant favorable alleles due to their linkage. This assumption is supported by the finding that heterotic QTL as well as genomic loci containing heterosis-correlated genes exhibit enrichments for centromeric regions (Schön et al. 2010, Larièpe et al. 2012, Thiemann et al. 2014). Furthermore, it was shown that linkage phases are highest in pericentromeric regions (Technow et al. 2014). It was shown that pericentromeric regions are rich in pseudogenes (Hall et al. 2006). The direct or read-through transcription of these pseudogenes results in sRNAs (Liu et al. 2014), which can be assumed to target other homologous target genes in trans.

The distribution of ha-sRNA to annotation types reveals 22 nt ha-sRNAs as putatively *trans* factors due to their mapping to genic loci along with repeat and/or intergenic loci. The 24 nt ha-sRNAs are less represented in overlapping annotations but in single annotations, especially regions annotated as intergenic or repeat. This suggests 24 nt ha-

sRNAs to more likely be acting primarily in *cis* at their locus of generation or in *trans* at TEs of the same family respectively.

He et al. (2013) suggested that intergenic regions are associated with DNA methylation and also suggested these epigenetically diverse regions to be key drivers of variation, stability, and activity of the hybrid genomes. The mentioned high non-collinearity and polymorph characteristic of intergenic regions (including TEs) in maize results in differences in the DNA sequences of regulatory regions up-/downstream of genes, and furthermore gives rise to epialleles through RdDM, finally resulting in transcriptional variation of affected genes (Lu et al. 2005, Song & Messing 2003, Stupar & Springer 2006). It was shown that in unstressed Arabidopsis thaliana plants, the rate of spontaneous DNA methylation gain or loss events is 1,000-fold higher than the genetic mutation rate (Becker et al. 2011), allowing the generation of new epialleles, and was hypothesized to thus might have a higher impact on evolution than genetic mutation (Matzke & Mosher 2014). These changes in DNA methylation are less frequent at loci producing siRNAs, suggesting RdDM to have a conserving effect for epiallele states (Becker et al. 2011). These epialleles can be stably transmitted to the progeny, and thus were suggested to represent an adaptive epigenetic Lamarckian inheritance (Matzke & Mosher 2014). After hybridization, interallelic RdDM results in gain and loss of parental epialleles in trans, depending on the expression levels of siRNAs in the parents exceeding or falling below certain expression thresholds required for the establishment or maintenance of an epigenetic state (Greaves et al. 2012, Chodavarapu et al. 2012). It was shown for LTR retrotransposons that silencing of active copies in one parent is triggered in *trans* by the silenced TEs in the other parent even if the expression of the TE is below the expression threshold of RdDM initiation (Marí-Ordóñez et al. 2013). The regulatory regions of genes can be affected by spreading of DNA-methylation by RdDM from adjacent TEs and as a consequence result in transcriptional changes (Slotkin & Martienssen 2007, Lisch 2009, Wei et al. 2014).

The relative increase of 24 nt ha-sRNAs adjacent to genes resembles results from previous studies (Gent *et al.* 2012, Wei *et al.* 2014). A recent study analyzing the DNA methylation in *Arabidopsis thaliana* revealed distinct methylation patterns for methylation contexts adjacent to genes. While CG and CHG context methylation dropped near genes, an opposing increase was shown for the CHH context (Gent *et al.* 2013). The congruence of 24 nt siRNAs and CHH methylation was shown by various studies (Law & Jacobsen 2010, Gent *et al.* 2013, Regulski *et al.* 2013). Gent *et al.* (2013) furthermore showed a positive correlation of gene expression and flanking CHH methylation, and suggested these "CHH islands" to act as insulators that prevent the spreading of heter-ochromatin into promoters and 3' regulatory regions.

22 nt ha-sRNAs from intergenic regions are primarily localized distant to genes. If they are actually located within genes, they are mostly found within introns. Chen *et al.* (2011) uncovered a mechanism generating siRNAs from introns that were able to form stem-loop structures called "sirtrons" similar to pre-miRNA but processing into 21, 22, and 24 nt siRNAs from their long stem regions. These sirtron-derived siRNAs were shown to mediate DNA methylation, and thus regulation of their host gene (Chen *et al.* 2011).

The 24 nt ha-sRNAs exhibit an opposite distribution to 22 nt ha-sRNAs relative to genes. The fraction of 24 nt ha-sRNAs is highest adjacent to genes and at exon-intron junctions. Recently, it was shown that CHG methylation at splice acceptor sites may inhibit RNA splicing (Regulski *et al.* 2013). Thus, the 24 nt ha-sRNAs with a trend towards these regions might act in splicing regulation. Wei *et al.* (2014) revealed that agricultural traits are affected by knockdown mutants of DCL3, which is involved in the generation of 24 nt siRNAs. The study revealed a reduction of TE-derived siRNAs that map in the 5', 3' or intronic region of genes (Wei *et al.* 2014). These results are supported by the finding of this study that ha-sRNAs represent the largest amount of sRNAs adjacent to genes and a large fraction of pos. ha-sRNAs. Similar to the study by Wei *et al.* (2014) a removal of these 24 nt pos. ha-sRNAs might have a negative impact on trait-associated genes.

The analysis for enrichment of ha-sRNAs to TEs reveals the 22 nt ha-sRNAs to be enriched for the LTR Gypsy retrotransposon super-family and its families cinful-zeon, gyma and nihep. This enrichment for LTR Gypsy retrotransposons explains the enrichment of predominantly neg. 22 nt ha-sRNAs in the pericentromeric regions, as LTR Gypsy are known to be predominantly located in the heterochromatic pericentromeres (Baucom *et al.* 2009). It was shown that LTR Gypsy retrotransposons generally exhibit a higher level of transcription in comparison to other TEs super-families (Vicient 2010). The overrepresentation of 22 nt siRNAs was shown by other studies, not only for LTR Gypsy but also for LTR Copia retrotransposons (Barber *et al.* 2012, He *et al.* 2013), which are primarily located in euchromatic regions in the chromosome arms (Baucom *et*

al. 2009). In this study, the LTR Copia retrotransposon families ji and leviathan are exceptions, showing an enrichment for pos. and neg. 22 nt ha-sRNAs respectively, although LTR Copia retrotransposons are primarily located within the euchromatic arms of the chromosomes (Mroczek & Dawe 2003). The repeat family ji was shown by Barber et al. (2012) to exhibit parental differences for 22 nt sRNAs between the two maize inbred lines B73 and Mo17. The neg. 22 ha-sRNA enrichment for cinful-zeon is also consistent with the findings of 22 nt sRNAs with differences between inbred lines by Barber et al. (2012) in contrast to the other LTR Gypsy repeat-families with an enrichment for neg. 22 nt ha-sRNAs, namely gyma and nihep, or the LTR Copia repeat-family leviathan. Thus, it can be shown that the enrichments for ha-sRNAs do not solely resemble the differences of any pairs of inbred lines. This suggests a specific function for the neg. 22 nt ha-sRNAs, and a thus resulting deviation of the enrichment patterns towards these specific ha-sRNAs. It was shown that in the maize MOP1 mutant plants 22 nt siRNAs were not affected, suggesting a second mechanism not including RDR2 to be involved in their generation (Nobuta et al. 2008, Barber et al. 2012). It was shown recently in Arabidopsis that 22 nt siRNAs, dependent on the RNA-dependent RNA polymerase RDR6, are involved in the establishment of silencing of active transposable elements, mainly belonging to the LTR Gypsy retrotransposon super-family (McCue et al. 2012, Nuthikattu et al. 2013). Ohtsu et al. (2007) suggested that the expression of retrotransposons in meristematic plant tissues provides a source for siRNAs that target genes in their untranslated regions and result in additional regulatory effects in dividing tissues. TEs were shown to be highly expressed in the shoot apical meristem, with the majority of transcripts being derived from LTR Gypsy retrotransposons (Vicient 2010). The hypothesis of gene targeting of TE sRNAs was confirmed for a small RNA from the Athila family of Arabidopsis thaliana LTR Gypsy retrotransposons, targeting the untranslated region of the gene UBP1b, resulting in inhibition of translation, and recently further candidates were identified (McCue et al. 2012, McCue et al. 2013). Interestingly, TEs exhibit a high level of sequence diversity even if their copy numbers are in the tens of thousands. This is caused by the accumulation of mutations if they were not both recently and perfectly transposed (Feschotte et al. 2002). This provides a source of sequence diversity for sRNA generated from these regions for the targeting of genes and variation in inbred lines having effects in the generation or alteration of epialleles in resulting hybrids. Interestingly, LTR Gypsy retrotransposons were shown to be overrepresented among retrotransposon families that result in spreading of DNA methylation

and H3K9 chromatin modifications, and thus affected the expression of neighboring genes (Eichten *et al.* 2012).

The 24 nt ha-sRNAs show enrichments for DNA transposons, except for Tc1/Mariner, and depletion for retrotransposons with the exception of the super-family LTR Unknown. This enrichment resembles results from other studies for 24 nt siRNAs (He *et al.* 2013). Interestingly, all DNA transposon super-families exhibit enrichments for both pos. and neg. 24 nt ha-sRNAs, with the exception of the super-family Pif/Harbinger which is only enriched for neg. ha-sRNAs. The DNA transposons super-families Pif/Harbinger and Mutator were shown to be located at median distance of 1-6 kbp from genes, while CACTA DNA transposons are exhibiting a median distance of 12-27 kbp to genes (Gent *et al.* 2013). Thus, for at least some of the DNA transposons super-families reveals an unspecific regulatory effect to adjacent genes. The effect of the regulation resulting in the grouping into pos. or neg. ha-sRNAs is most probably not solely dependent on the ha-sRNA but more likely on the effect of affected genes on the establishment of heterosis, which can be either favorable or detrimental.

Deviating from all other TE families, the LTR Unknown family ipiki is enriched for all sRNA lengths of both pos. and neg. ha-sRNAs. The TEs from the LTR Unknown retrotransposon family ipiki were assumed to be located at centromeres (Baucom *et al.* 2009) and more recently found to be enriched at the short arm of maize chromosome 6 (Dukowic-Schulze *et al.* 2014). The mentioned enrichment at the short arm of chromosome 6 is visible in the region from 10 to 15 Mbp for the distribution of all sRNAs from the 21 inbred lines and three hybrids as well as the ha-sRNAs. The enrichment of sRNAs from the LTR Unknown retrotransposon family ipiki throughout the whole analyzed spectrum of sRNA lengths suggests an unspecific generation of those sRNAs. This sRNA length distribution most likely reflects the degradation of ipiki transcripts. The lack of any length pattern specific for any known siRNA class leads to the assumption that the ipiki-derived ha-sRNAs more likely reflect an effect of hybridization in relation to heterosis, rather than sRNAs beeing functionally involved in heterosis formation.

The expression patterns of differentially expressed sRNAs in general as well as hasRNAs in particular tend towards a non-additive low-parent-like expression. This result is consistent with other studies (Groszmann *et al.* 2011, Barber *et al.* 2012, Li *et al.* 2012). Li *et al.* (2012) observed the low-parent-like expression especially for sRNA clusters co-localized with protein-coding genes, in contrast to TE loci being additively expressed. The majority of the ha-sRNAs is not exclusively associated with genes. In fact, it is possible that the expression patterns of ha-sRNA are in general altered after hybridization, and thus result in low-parent-like expression. This effect might not have been observed by the mentioned studies from Groszmann *et al.* (2011), Barber *et al.* (2012), and Li *et al.* (2012), since they consider differentially expressed clusters of siRNAs instead of distinct sRNA sequences. These clusters most probably mask the non-additive expression patterns of contained ha-sRNAs due to the aggregation of sRNAs of different expression patterns that are mapping at hundreds to thousands of loci in the genome.

4.4.4 Analysis of effects on the transcriptome mediated by ha-sRNAs

The target prediction for the ha-sRNAs resulting in a large number of targets gives rise to the assumption that at least some of the ha-sRNAs might act in post-transcriptional gene silencing or inhibition of translation. The high ratio of targets per sRNAs for the 18 nt and 19 nt ha-sRNAs might be caused by the short sRNA length, and thus resulting in higher statistical probability of potential nearly complementary gene transcripts. The high number of ~4 targets per sRNA for the 20 nt to 22 nt ha-sRNAs might reflect the observed regulatory effect of sRNAs with equal sRNA lengths known to be involved in PTGS (Martínez de Alba et al. 2013). This assumption is supported by significant and strong linear correlations for a subset of the ha-sRNAs with microarray transcriptome expression data. Both positive and negative correlations were found for the ha-sRNAs and putative target transcripts. ha-sRNAs effecting PTGS are expected to exhibit a negative correlation with their target gene expression. In contrast, a positive correlation might point towards the generation of secondary siRNAs from their putative target loci. It should be noted that this analysis was not expected to elucidate the most relevant hasRNAs target gene relations due to the shortcoming of the analysis to rely on linear relations. Nevertheless, the analysis allowed an insight in a subset of putatively ha-sRNA regulated genes and revealed the unexpected enrichment of 22 nt ha-sRNAs as well as the trend towards an on average four times higher variety of genes affected by pos. hasRNAs than shown for ha-sRNAs, although the number of pos. and neg. ha-sRNA subsets were of similar size.

The analysis of the ha-sRNAs identified in the breeding population of the University of Hohenheim in the genetically divergent inbred lines B73 and Mo17 revealed 42,65 % of the ha-sRNAs to be present, with only 109 being differentially expressed. Most strikingly, the 22 nt ha-sRNAs exhibit a larger fraction of ha-sRNAs differentially expressed between B73 and Mo17 than for the 21 inbred lines from the University of Hohenheim, suggesting a conservation of their regulatory function. This trend is even stronger for the ha-sRNAs found to be differentially expressed between the lines B73 or Mo17.

The analysis of the enrichment of ha-sRNA adjacent to or at differentially expressed genes between B73 and Mo17 revealed the predominant trend of 22 nt ha-sRNAs to be enriched within genes, in contrast to the 24 nt ha-sRNA with enrichments shown in similar amount 1 kbp upstream/downstream as well as in the gene. Considering the finding of predominant distribution within introns rather than exons, this suggests 22 nt ha-sRNAs to act by silencing of genes by directing DNA methylation of their introns. This is in partial consistent with the finding by Chen *et al.* (2011) who found siRNAs from introns to mediate DNA methylation of their host genes. The results of this study suggest that it is more likely that the 22 nt ha-sRNAs act in *trans* rather than in *cis* on these genes. A possible link to this contradiction with the results by Chen *et al.* (2011) might propose the 22 nt ha-sRNAs to act in *trans*, but initiate a mechanism which is then propagated by *cis* acting mechanisms.

4.4.5 ha-sRNAs hold the potential for the prediction of hybrid traits and implications for future plant breeding

Hybrid breeding gained in impact to crop improvement since its proposal for maize (Shull 1908, Duvick 1999). The importance of hybrid breeding was even increased by the introduction of the doubled haploid technology enabling the fast generation of homozygous inbred lines (Technow *et al.* 2014). Although the separation of germplasm into divergent heterotic pools resulted in the generation of hybrids yielding high heterosis (Melchinger & Gumber 1998), the selection of inbred line pairs resulting in highest possible heterosis is pivotal. It is not possible to perform all potential testcrosses of any inbred pairs from two heterotic pools (Bernardo 1996). Thus, to optimally tract this selection problem and reduce the number of field tests for the determination of the

inbred combination yielding in the best hybrids, genomic prediction approaches were developed (Technow *et al.* 2014).

Currently, plant breeding programs rely on DNA-based marker data (e.g. AFLP, SNP), although it was shown that heterosis cannot be explained solely based on the genetic distance (Melchinger 1999) and even plants with nearly identical genome sequence but different epigenetic status exhibit high heterosis (Shen *et al.* 2012).

It could be shown that ha-sRNAs are highly predictive for heterosis (MPH for GY) as well as hybrid performance for grain yield (HP for GY). The average prediction accuracy of 0.88 for Type-2 prediction of MPH for GY exceeds previous results based on AFLP, and transcriptome data (Schrag *et al.* 2007, Frisch *et al.* 2010), which were performed for the identical breeding factorial as in this study. The Type-2 prediction of HP for GY resulted in higher accuracy in comparison to predictions based on general combining ability (GCA) or AFLP data and slightly higher accuracy than transcriptome-based prediction for all analyses based on the identical breeding experiment (Schrag *et al.* 2007, Frisch *et al.* 2010).

It could be shown that intraspecific combinations of genetically divergent genomes are not in every case resulting in higher heterosis. The antagonistic acting of pos. and neg. ha-s RNAs with specific characteristics for the neg. 22 nt ha-sRNAs, regarding to their genomic localization and high copy number of loci of generation/acting, holds implications for the optimization of inbred line selection in hybrid plant breeding programs. These neg. 22 nt ha-sRNAs are enriched in the pericentromeric regions, which are known to exhibit low recombination rates, and thus are intractable for breeders. Furthermore, it was shown that heterosis-associated genes in the identical breeding population are located in pericentromeric regions (Thiemann et al. 2014), suggesting a repulsion phase linkage of the neg. 22 nt ha-sRNA generation loci with these genes. The pos. ha-sRNAs do not exhibit enrichments for specific regions of the reference genome, and thus might be less conserved and stable in comparison to the neg. 22 nt ha-sRNAs. Nevertheless, the results suggest that an optimization towards inbred pairs with largest amounts of differentially expressed pos. ha-sRNAs and a minimization of differentially expressed neg. ha-sRNAs will result in the optimal trait value (e.g. MPH for GY) of interest.

The prediction of heterosis and hybrid performance with high precision highlights small RNAs and epigenetics in general as crucial selection markers for future crop improve-

ment. The integration of knowledge about epigenetics of inbred parents and hybridization induced changes of epigenetic states into breeding programs should enable plant breeders to access currently missing heritability to increase heterosis.

5. Perspective

The identification of ha-sRNAs in a factorial of a breeding population of the University of Hohenheim, shown in this study, suggests further analyses for the analysis of ha-sRNAs. It is of high interest, to ascertain to what extent the ha-sRNAs are representative in other breeding factorials. Furthermore, it should be analyzed if the composition of ha-sRNAs regarding the ratio of pos. and neg. ha-sRNAs as well as the ha-sRNA length distribution exhibits the same trends in a larger set of inbred lines and resulting hybrids.

Similar to the analysis of ha-sRNAs in the dataset from Regulski *et al.* (2013), a reanalysis of a dataset generated by He *et al.* (2013) should be performed. This would allow analyzing the function of ha-sRNAs from this study regarding DNA methylation and histone modification data, and thus might generate a more detailed insight in the regulatory effects guided by ha-sRNAs.

As the annotation of ha-sRNAs in this study relies on the mapping of ha-sRNAs to the B73 reference genome and a high level of sequence variation was shown between maize inbred lines, the re-sequencing of the inbred lines should allow a more detailed analysis of regulatory targets of ha-sRNAs in the divergent inbred lines.

Besides continuing and more detailed analyses in maize, the identification of ha-sRNAs should similarly be performed in plants of differing sequence variation between inbred lines of monocot and dicot plant species as wheat, *Arabidopsis*, and rapeseed to confirm if ha-sRNAs are related to sequence variation and fixation in regions of recombination suppression.

It should be furthermore analyzed if the normalization of sequence expression data based on the bootstrap-based method performs equally to the quantile normalization and both methods should be evaluated via experimental validation.

VII. Abstract

Heterosis, a phenomenon resulting in the superior performance of heterozygous hybrids in comparison to their homozygous parental inbred lines, is extensively exploited in agriculture, but the molecular basis is still unknown. It was hypothesized that epigenetics and small RNAs are components in the establishment of heterosis. Various studies showed differences in small RNA expression between two inbred lines as well as their hybrid offspring, but whether small RNAs are involved in the formation of heterosis remained unanswered.

In this study, bioinformatical analyses of small RNA sequencing libraries of hybrids containing an inducible gene silencing suppressor, which affects small RNAs, were performed to elucidate a functional participation in the establishment of heterosis. The analysis revealed a significant increase of heterosis due to the reduction of the small RNA population complexity.

A second experiment was performed to analyze and associate the contribution of sRNAs in a multiplicity of parental genotypes from a maize breeding population with their hybrid traits. The analysis revealed antagonistically acting heterosis associated small RNAs, with a predominant negative component in the formation of heterosis. The heterosis associated small RNAs were found to exhibit distinct localization throughout the maize genome and relative to genes, allowing the assumption of putative mechanisms of acting. The heterosis associated small RNAs in the inbred parents were shown to be highly predictive for the heterosis observed in their hybrids. The findings of this study reveal small RNAs as an important component in the establishment of heterosis in maize uncovering currently hidden potential in hybrid breeding.

VIII. Zusammenfassung

Heterosis beschreibt das Phänomen der überlegenen Leistungsfähigkeit von heterozygten Hybriden im Vergleich zu deren homozygoten parentalen Inzuchtlinien. Trotz der langen und erfolgreichen Nutzung der Hybridzucht in der Landwirtschaft konnte die molekulare Grundlage der Heterosis noch nicht aufgeklärt werden. Es wurde vermutet, dass epigenetische Mechanismen und kurze RNAs eine Komponente in der Entstehung der Heterosis sind. Eine Vielzahl jüngster Studien zeigten Unterschiede in der Expression von kurzen RNAs sowohl zwischen Inzuchtlinien als auch zu deren Hybridnachkommen. Ob diese Unterschiede eine Beteiligung in der Ausbildung der Heterosis bedeuten, blieb jedoch unbeantwortet.

In dieser Arbeit wurde über bioinformatische Analysen von kurzen RNA-Tiefensequenzierungsdaten von Hybriden, welche einen induzierbaren viralen Gene Silencing Suppressor mit einer Wirkung auf kurze RNAs enthalten, eine funktionelle Beteiligung der kurzen RNA an der Entstehung von Heterosis untersucht. Die Untersuchung deckte eine signifikante Zunahme der Heterosis aufgrund der Reduktion der Populationskomplexität von kurzen RNAs auf.

In einem zweiten Experiment wurde die Beteiligung von kurzen RNAs an der Entstehung von Heterosis in einer Vielzahl von elterlichen Genotypen einer Mais-Zuchtpopulation über die Assoziation mit deren resultierenden Hybridmerkmalen untersucht. Die Analyse deckte antagonistisch wirkende heterosis-assoziierte kurze RNAs mit einem überwiegend negativen Anteil in der Ausbildung von Heterosis auf. Die heterosis-assoziierten kurzen RNAs weisen unterschiedliche Lokalisierungen im Maisgenom und relativ zu Genen auf, was einen Rückschluss auf mögliche Wirkmechanismen erlaubt. Die heterosis-assoziierten kurzen RNAs in den elterlichen Linien weisen eine hohe Vorhersagekraft für die in den Hybriden beobachtete Heterosis auf. Die Ergebnisse dieser Arbeit decken kurze RNAs als eine wichtige Komponente in der Entstehung von Heterosis in Mais auf und deuten auf ein bisher verborgenes Potential in der Hybridzucht hin.

IX. References

Altschul, S.F., Madden T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25, 3389-3402

Axtell, M.J., Bowman, J.L. (2008) Evolution of microRNAs and their targets: Are all microRNAs biologically relevant? *Trends in Plant Science* 13 (7), 343-349

Axtell, M.J. (2013) Classification and Comparison of Small RNAs from Plants. Annual Review of Plant Biology 64, 137-159

Azevedo, J., Garcia, D., Pontier, D., Ohnesorge, S., Yu, A., Garcia, S., Braun, L., Bergdoll, M., Hakimi, M.A., Lagrange, T., Voinnet, O. (2010) Argonaute quenching and global changes in Dicer homeostasis caused by a pathogen-encoded GW repeat protein. *Genes & Development* 24, 904-915

Barber, W.T., Zhang, W., Win, H., Varala, K.K., Dorweiler, J.E., Hudson, M.E., Moose, S.P. (2012) Repeat associated small RNAs vary among parents and following hybridization in maize. *Proceedings of the National Academy of Sciences of the United States of America* 109 (26), 10444-10449

Bartel, D.P. (2004) microRNAs. Genomics, Biogenesis, Mechanism, and Function. *Cell* 116, 281-297

Barthelson, R., McFarlin, A.J., Rounsley, S.D., Young, S. (2011) Plantagora: Modeling Whole Genome Sequencing and Assembly of Plant Genomes. *PloS ONE* 6 (12), e28436

Barros, E., Lezar, S., Anttonen, M.J., van Dijk, J.P., Röhlig, R.M., Kok, E.J., Engel, K.H. (2010) Comparison of two GM maize varieties with near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. *Plant Biotechnology Journal* 8 (4), 436-451

Baucom, R.S., Estill, J.C., Chaparro, C., Upshaw, N., Jogi, A., Deragon, J.M., Westerman, R.P., SanMiguel, P.J., Bennetzen, J.L. (2009) Exceptional Diversity, Non-Random Distribution, and Rapid Evolution of Retroelements in the B73 Maize Genome. *PLoS Genetics* 5 (11), e1000732

Baulcombe, D.C. (2004) RNA silencing in plants. Nature 431, 356-363

Baumberger, N., Baulcombe, D.C. (2005) *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proceedings of the National Academy of Sciences of the United States of America* 102 (33), 11928-11933

Baumberger, N., Tsai, C.H., Lie, M., Havecker, E., Baulcombe, D.C. (2007) The Polerovirus silencing suppressor P0 targets ARGONAUTE proteins for degradation. *Current Biology* 17 (18), 1609-1614

Becker, C., Hagmann, J., Müller, J., Koenig, D., Stegle, O., Borgwardt, K., Weigel, D. (2011) Spontaneous epigenetic variation in *Arabidopsis thaliana* methylome. *Nature* 480, 245-249

Benjamini, Y., Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1), 289-300

Bernardo, R. (1996) Best linear unbiased prediction of maize single-cross performance. *Crop Science* 36, 50-56

Bernstein, E., Caudi, A.A., Hammond, S.M., Hannon, G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363-366

Bertie, A.J. (2004) http://www.jsc.nildram.co.uk

Birchler, J.A., Auger, D.L., Riddle, N.C. (2003) In Search of the Molecular Basis of Heterosis. *The Plant Cell* 15, 2236-2239

Birchler, J.A., Yao, H., Chudalayandi, S., Vaiman, D., Veitia, R.A. (2010) Heterosis. *The Plant Cell* 22, 2105-2112

Bolstad, B.M., Irizarry, R.A., Åstrand, M., Speed, T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19 (2), 185-193

Bonnet, E., van de Peer, Y., Rouzé, P. (2006) The small RNA world of plants. *New Phytologist* 171, 451-468

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., Zhu, J.K. (2005) Endogenous siRNAs Derived from a Pair of Natural *cis*-Antisense Transcripts Regulate Salt Tolerance in *Arabidopsis*. *Cell* 123, 1279-1291

Bortolamiol, D., Pazhouhandeh, M., Marrocco, k., Genschik, P., Ziegler-Graff, V. (2007) The Polerovirus F box protein P0 targets ARGONAUTE1 to suppress RNA silencing. *Current Biology* 17 (18), 1614-1621

Bosch, M., Mayer, C.D., Cookson, A., Donnison, I.S. (2011) Identification of genes involved in cell wall biogenesis in grasses by differential gene expression profiling of elongating and nonelongating maize internodes. *Journal of Experimental Biology* 62 (10), 3545-3561

Brodersen, P., Voinnet, O. (2006) The diversity of RNA silencing pathways in plants. *TRENDS in Genetics* 22 (5), 268-280

Bruce, A.B. (1910) The Mendelian theory of heredity and the augmentation of vigor. *Science* 32, 627-628

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421

Campenot, M.K., Zhang, G., Cutler, A.J., Cass, D.D. (1992) *Zea mays* Embryo Sacs in Culture I. Plant Generation form 1 Day After Pollination Embryos. *American Journal for Botany*, 1386-1373

Campos-Bermudez, V.A., Fauguel, C.M., Tronconi, M.A., Casati, P., Presello, D.A., Andreo, C.S. (2012) Transcriptional and metabolic changes associated to the infection by *Fusarium verticillioides* in maize inbreds with contrasting ear rot resistance. *PLoS ONE* 8 (4): e61580

Cardwell, V.B. (1982) Fifty Years of Minnesota Corn Production: Sources of Yield Increase. *Agronomy Journal* 74 (6), 984-990

Carrington, J.C., Ambros, V. (2003) Role of microRNAs in Plant and Animal Development. *Science* 301, 336-338

Castel, S.E., Martienssen, R.A. (2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nature Reviews Genetics* 14, 100-112

Charlesworth, D., Willis, J.H. (2009) The genetics of inbreeding depression. *Nature Reviews Genetics* 10, 783-796

Chellappan P., Xia, J., Zhou, X., Gao, S., Zhang, X., Coutino, G., Vazquez, F., Zhang, W, Jin, H. (2010) siRNAs from miRNA sites mediate DNA methylation of target genes. *Nucleic Acids Research* 38 (20), 6883-6894

Chen, X. (2009) Small RNAs and Their Roles in Plant Development. *Annual Review of Cell and Developmental Biology* 35, 21-44

Chen, F., He, G., he, H., Chen, w., Zhu, X., Liang, M., Chen, L., Deng, X.W. (2010) Expression analysis of miRNAs and highly-expressed small RNAs in two rice subspecies and their reciprocal hybrids. *Journal of Integrative Plant Biology* 52 (11), 971-980

Chen, D., Meng, Y., Yuan, C., Bai, L., Huang, D., Lv, S., Wu, P., Chen, L.L., Chen, M. (2011) Plant siRNAs from introns mediate DNA methylation of host genes. *RNA* 17 (6), 1012-1024

Cheres, M.T., Miller, J.F., Crane, J.M., Knapp, S.J. (2000) Genetic distance as a predictor of heterosis and hybrid performance within and between heterotic groups in sunflower. *Theoretical and Applied Genetics* 100, 889-894

Chiu, M.H., Chen, I.H, Baulcombe, D.C., Tsai, C.H. (2010) The silencing suppressor P25 of *Potato virus X* interacts with Argonaute1 and mediates its degradation through the proteasome pathway. *Molecular Plant Pathology* 11 (5), 641-649

Chodavarapu, R.K., Feng, S., Ding, B., Simon, S.A., Lopez, D., Jia, Y., Wang, G.L., Meyers, B.C., Jacobsen, S.E., Pellegrini, M. (2012) Transcriptome and methylome interactions in rice hybrids. *Proceedings of the National Academy of Sciences of the United States of America* 109 (30), 12040-12045

Conesa, A., Götz, S. (2008) Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *International Journal of Plant Genomics* 2008, 1-13

Coruh, C., Shahid, S., Axtell, M.J. (2014) Seeing the forest for the trees: annotating small RNA producing genes in plants. *Current Opinion in Plant Biology* 18, 87-95

Covshoff, S., Majeran, W., Liu, P., Kolkman, J.M., van Wijk, K.J., Brutnell, T.P. (2008) Deregulation of maize C4 photosynthetic development in a mesophyll cell-defective mutant. *Plant Physiology* 164 (4), 1469-1481

Crow, J.F. (1948) Alternative Hypotheses of Hybrid Vigor. Genetics 33, 477-487

Crow, J.F. (1998) 90 Years Ago: The Beginning of Hybrid Maize. Genetics (148), 923-928

Cuellar, W.J., Kreuze, J.F., Rajamäki, M.L., Cruzado, K.R., Untiveros, M., Valkonen, J.P.T. (2009) Elimination of antiviral defense by viral RNAse III. *Proceedings of the National Academy of Sciences of the United States of America* 106 (25), 10354-10358

Cuperus, J.T., Fahlgren, N., Carrington, J.C. (2011) Evolution and Functional Diversification of MIRNA Genes. *The Plant Cell* 23, 431-442

Czech, B., Hannon, G. (2011) Small RNA sorting: matchmaking for Argonautes. *Nature Reviews Genetics* 12 (1), 19-31

Dai, X., Zhao, P.X. (2011) psRNATarget: a plant small RNA target analysis server. *Nucleic Ac-ids Research* 39 (suppl 2), W155-W159

Darwin, C. (1876) The effects of cross and self fertilization in the vegetable kingdom. London: John Murray

Daxinger, L., Kanno, T., Bucher, E., van der Winden, J., Naumann, U., Matzke, A.J.M., Matzke, M. (2009) A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. *The EMBO Journal* 28, 48-57

Dillies, M.A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., Guenec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Le Crom, S., Guedj, M., Jaffrézic, F., The French StatOmique Consortium (2012) A comprehensive evaluation of normalization methods for Illumina highthroughput RNA sequencing data analysis. *Briefings in Bioinformatics* 14 (6), 671-683

Ding, D., Wang, Y., Han, M., Fu, Z., Li, W., Liu, Z., Hu, Y., Tang, J. (2012) MicroRNA Transcriptomic Analysis of Heterosis during Maize Seed Germination. *PloS ONE* 7 (6), e39578

Dorsett, Y., Tuschl, T. (2004) siRNAs: applications in functional genomics and potential as therapeutics. *Nature Reviews Drugs Discovery* 3 (4), 318-329

Dukowic-Schulze, S., Sundarajan, A., Mudge, J., Ramaraj, T., farmer, A.D., Wang, M., sun, Q., Pillardy, J., Kianian, S., Retzel, E.F., Pawlowski, W.P., Chen, C. (2014) The transcriptome landscape of early maize meiosis. *BMC Plant Biology* 14:118

Duvick, D.N. (1999) Heterosis: Feeding people and protecting natural resources. In: CIMMYT (1997) *Book of Abstracts: The Genetics and Exploitation of Heterosis in Crops; An International Symposium.* Mexico, D.F., Mexico

Duvick, D.N., Smith, J.S.C., Cooper, M. (2004) Long-term selection in a commercial hybrid maize breeding program. *In: Janick, J. (editor) Plant breeding reviews 24 (2). Long term selection: crops, animals, and bacteria.* Wiley, New York (NY, USA), 109-151.

Duvick, D.N. (2005) Genetic progress in yield of United States maize (*Zea mays* L.). *Maydica* 50, 193-202

East, E.M. (1936) Heterosis. Genetics 21, 375-397

Eichten, S.R., Ellis, N.A., Makarevitch, I., Yeh, C.T., Gent, J.I., Guo, I., McGinnis, K.M., Zhang, X., Schnable, P.S., Vaughn, M.W., Dawe, R.K., Springer, N.M. (2012) Spreading of Heterochromatin Is Limited to Specific Families of Maize Retrotransposons. *PloS Genetics* 8 (12), e1003127

Eichten, S.R., Briskine, R., Song, J., Li, Q., Swanson-Wagner, R., Hermanson, P.J., Waters, A.J., Starr, E., West, P.T., Tiffin, P., Myers, C.L., Vaughn, M.W., Springer, N.M. (2013) Epigenetic and Genetic Influences on DNA Methylation Variation in Maize Populations. *The Plant Cell* 25 (8), 2783-2797

Engelberth, J. (2013a) http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE47982

Engelberth, J. (2013b) http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50981

Fei, Q., Xia, R., Meyers, B.C. (2013) Phased, Secondary, Small Interfering RNAs in Posttranscriptional Regulatory Networks. *The Plant Cell* 25, 2400-2415

Fernandes, J., Morrow, D.J., Casati, P., Walbot, V. (2008) Distinctive transcriptome responses to adverse environmental conditions in *Zea mays* L. *Plant Biotechnology Journal* 6 (8), 782-798

Feschotte, C, Jiang, N., Wessler, S.R. (2002) Plant Transposable Elements: Where Genetics meets Genomics. *Nature Reviews Genetics* 3, 329-341

Finnegan, E.J., Matzke, M.A. (2003) The small RNA world. Cell Science at a Glance 116, 4689-4693

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

Flint-Garcia, S.A., Buckler, E.S., Tiffin, P., Ersoz, E., Springer, N.M. (2009) Heterosis Is Prevalent for Multiple traits in Diverse Maize Germplasm. *PLoS ONE* 4 (10), e7433

Frisch, M., Thiemann, A., Fu, J., Schrag, T.A., Scholten, S., Melchinger, A.E. (2010) Transcriptome-based distance measures for grouping of germplasm an prediction of hybrid performance in maize. *Theoretical and Applied Genetics* 120 (2), 441-450

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., Paul, C.L. (1992) A genomic sequencing protocol that yields a positive display of 5methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America*, 89 (5), 1827-1831

Fu, **J.**, **Falke**, **C.**, **Thiemann**, **A.**, **Schrag**, **T.A.**, **Scholten**, **S.**, **Melchinger**, **A.E.** (2012) Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize. *Theoretical and Applied Genetics* 124, 825-833

Fujimoto, R., Taylor, J.M., Shirasawa, S., Peacock, W.J., Dennis, E.S. (2011) Heterosis of Arabidopsis hybrids between C24 and Col is associated with increased photosynthesis capacity. *Proceedings of the National Academy of Sciences of the United States of America* 109 (18), 7109-7114

Gardiner, J.M., Buell, C.R., Elumalai, R., Galbraith, D.W., Henderson, D.A., Iniguez, A.L, Kaeppler, S.M., Kim, J.J., Smith, A., Zheng, L., Chandler, V.L. (2005) Design, Production, and Utilization of Long Oligonucleotide Microarrays for Expression Analysis in Maize. *Maydica* 50, 425-435

Gazzani, S., Lawrenson, T., Woodward, C., Headon, D., Sablowski, R. (2004) A Link Between mRNA Turnover and RNA Interference in *Arabidopsis. Science* 306, 1046-1048

Gent, J.I., Dong, Y., Jiang, J., Dawe, R.K. (2012) Strong epigenetic similarity between maize centromeric and pericentromeric regions at the level of small RNAs, DNA methylation and H3 chromatin modifications. *Nucleic Acids Research* 40 (4), 1550-1560

Gent, J.I., Ellis, N.A., Guo, L., Harkess, A.E., Yao, Y., Zhang, X., Dawe, R.K. (2013) CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. *Genome Research* 23 (4), 628-637

Ghildiyal, M., Zamore, P.D. (2009) Small silencing sRNAs: an expanding universe. *Nature Reviews Genetics* 10, 94-108

Glick, E., Zrachya, A., Levy, Y., Mett, A., Gidoni, D., Belausov, E., Citovski, V., Gafni, Y. (2008) Interaction with host SGS3 is required for suppression of RNA silencing by the tomato yellow leaf curl virus V2 protein. *Proceedings of the National Academy of Sciences of the United States of America* 105 (1), 157-161

Goodnight, C.J. (1999) Epistasis and heterosis. In: Coors, J.G., Pandey, S., eds. Genetic and Exploitation of Heterosis in Crops. American Society of Agronomy, Crop Science Society of America, 59-67

Gore, M.A., Chia, J.M., Elshire, R.J., Sun, Q., Ersoz, E.S., Hurwitz, B.L., Peiffer, J.A., McMullen, M.D., Grills, G.S., Ross-Ibarra, J., Ware, D.H., Buckler, E.S. (2009) A First-Generation Haplotype Map of Maize. *Science* 326, 1115-1117

Greaves, I.K., Groszmann, M., Ying, H., Taylor, J.M., Peacock, W.J., Dennis, E.S. (2012) Trans Chromosomal Methylation in *Arabidopsis* hybrids. *Proceedings of the National Academy* of Sciences of the United States of America 109 (9), 3570-3575 **Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., Enright, A.J.** (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research* 34 (suppl 1), D140-D144

Groszmann, M., Greaves, I.K., Albertyn, Z.I., Scofield, G.N., Peacock, W.J., Dennis, E.S. (2011) Changes in 24-nt siRNA levels in *Arabidopsis* hybrids suggest an epigenetic contribution to hybrid vigor. *Proceedings of the National Academy of Sciences of the United States of America* 108 (6), 2617-2622

Groszmann, M., Greaves, I.K., Fujimoto, R., Peacock, W.J., Dennis, E.S. (2013) The role of epigenetics in hybrid vigour. *Trends in Genetics* 29 (12), 684-690

Guo, M., Rupe, M.A., Yang, X., Crasta, O., Zinselmeier, C., Smith, O.S., Bowen, B. (2006) Genome-wide transcript analysis of maize hybrids: allelic additive gene expression and yield heterosis. *Theoretical and Applied Genetics* 113, 831-845

Guo, H., Song, X., Xie, C., Huo, Y., Zhang, F., Chen, X., Geng, Y., Fang, R. (2013) *Rice yellow stunt rhabdovirus* Protein 6 Suppresses Systemic RNA Silencing by Blocking RDR6-Mediated Secondary siRNA Synthesis. *Molecular Plant-Microbe Interactions* 26 (8), 927-936

Ha, M., Lu, J., Tian, L., Ramachandran, V., Kasschau, K.D., Chapman, E.J., Carrington, J.C., Chen, X., Wang, X.J, Chen, Z.J (2009) Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proceedings of the National Academy of Sciences of the United States of America* 106 (42), 17835-17840

Haas, G., Azevedo, J., Moissiard, G., Geldenreich, A., Himber, C., Bureau, M., Fukuhara, T., Keller, M., Voinnet, O. (2008) Nuclear import of CaMV P6 is required for infection and suppression of the RNA silencing factor DRB4. *The EMBO Journal* 27, 2102-2112

Hall, A.E., Kettler, G.C., Preuss, D. (2006) Dynamic evolution at pericentromeres. *Genome Research* 16, 355-364

Hamilton, A., Voinnet, O., Chappell, L., Baulcombe, D. (2002) Two classes of short interfering RNA in RNA silencing. *The EMBO Journal* 21 (17), 4671-4679

Hansen, S., Horvath, D., Jarachi, Y., Carlson, G., Clay, D. (2011) http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33494

Hansey, C.N., Vaillancourt, B., Sekhon, R.S., de Leon, N., Kaeppler, S.M., Buell, C.R. (2012) Maize (*Zea mays* L.) Genome Diversity as Revealed by RNA-Sequencing. *PLoS ONE* 7 (3): e33071

Hayano-Kanashiro, C., Calderón-Vásquez, C., Ibarra-Laclette, E., Herrera-Estrella, L., Simpson, J. (2009) Analysis of gene expression and physiological responses in three Mexican maize landraces under drought stress and recovery irrigation. *PLoS ONE* 4 (10): e7531

He, G., Zhu, X., Elling, A.A., Chen, L., Wang, X., Guo, L., Liang, M., He, H., Zhang, H., Chen, F., Qi, Y., Chen, R., Deng, X.W. (2010) Global Epigenetic and Transcriptional Trends among Two Rice Subspecies and Their Reciprocal Hybrids. *The Plant Cell* 22, 17-33

He, G., Chen, B., Wang, X., Li, X., Li, J., He, H., Yang, M., Lu, L., Qi, Y., Wang, X., Deng, X. W. (2013) Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids. *Genome Biology* 14 (6): R57

Herr, A.J., Jensen, M.B., Dalmay, T., Baulcombe, D.C. (2005) RNA Polymerase IV Directs Silencing of Endogenous DNA. *Science* 308, 118-120

Hill, W.G., Robertson, A. (1966) The effect of linkage on limits to artificial selection. *Genetical Research* 8 (3), 269-294

Hirsch, C.N., Foerster, J.M., Johnson, J.M., Sekhon, R.S., Muttoni, G., Vaillancourt, B., Peñagaricano, F., Lindquist, E., Pedraza, M.A., Barry, K., de Leon, N., Kaeppler, S.M., Buell, C.R. (2014) Insights into the Maize Pan-Genome and Pan-Transcriptome. *The Plant Cell* 26 (1), 121-135

Hochholdinger, F., Hoecker, N. (2007) Towards the molecular basis of heterosis. *TRENDS in Plant Science* 12 (9), 427-432

Hoecker, N., Lamkemeyer, T., Sarholz, B., Paschold, A., Fladerer, C., Madlung J., Wurster, K., Stahl, M., Piepho, H.P., Nordheim, A., Hochholdinger, F. (2008) Analysis of nonadditive protein accumulation in young primary roots of a maize (*Zea mays* L.) F(1)-hybrid compared to its parental inbred lines. *Proteomics* 8 (18), 3882-3894

Holding, D.R., Hunter, B.G., Chung, T., Gibbon, B.C., Ford, C.F., Bharti, A.K., Messing, J., Hamaker, B.R., Larkings, B.A. (2008) Genetic Analysis of opaque 2 Modifier Loci in Quality Protein Maize. *Theoretical and Applied Genetics* 117 (2), 157-170

Hollister, J.D., Gaut, B.S. (2009) Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research* 19 (8), 1419-1428

Husson, F., Josse, J., Lê, S., Mazet, J. (2007) FactoMineR: Factor Analysis and Data Mining with R. R package version 1.26, URL http://CRAN.R-project.org/package=FactoMineR

Ihaka, R., Gentleman, R. (1996) R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5 (3), 299-314

Ito, H. (2012) Small RNAs and transposon silencing in plants. *Genes & genetic systems* 88 (1), 3-7

Jin, H., Vacic, V., Girke, T., Lonardi, S., Zhu, J.K. (2008) Small RNAs and the regulation of cis-natural antisense transcripts in *Arabidopsis*. *BMC Molecular Biology* 9:6

Johnson, E.T., Dowd, P.F., Liu, Z.L., Musser, R.O. (2011) Comparative transcription profiling analyses of maize reveals candidate defensive genes for seedling resistance against corn earworm. *Molecular Genetics and Genomics* 285 (6), 517-525

Johnson, E.T., Dowd, P.F. (2012)

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40107

Jones-Rhoades, M.W., Bartel, D.P., Bartel, B. (2006) microRNAs and Their Regulatory Roles in Plants. *Annual Review of Plant Biology* 57, 19-53

Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Carrington, J.C. (2007) Genome-Wide Profiling and Analysis of *Arabidopsis* siRNAs. *PLOS Biology* 5 (3), 479-493

Kenan-Eichler, M., Leshkowitz, D., Tal, L., Noor, E., Melamed-Bessudo, C., Feldman, M., Levy, A.A. (2011) Wheat Hybridization and Polyploidization Results in Deregulation of Small RNAs. *Genetics* 188 (2), 263-272

Krywinski, M., Schein, J.E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A. (2009) Circos: An information aesthetic for comparative genomics. *Genome Research* 19, 1639-1645

Kurihara, Y., Watanabe, Y. (2004) *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proceedings of the National Academy of Sciences of the United States of America* 101 (38), 12753-12758 Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10, R25

Larièpe, A., Mangin, B., Jasson, S., Combes, V., Dumas, F., Jamin, P., Lariagon, C., Jolivot, D., Madur, D., Fiévet, J., Gallais, A., Dubreuil, P., Charcosset, A., Moreau, L. (2012) The Genetic Basis of Heterosis: Multiparental Quantitative Trait Loci Mapping Reveals Contrasted Levels of Apparent Overdominance Among Traits of Agronomical Interest in Maize (*Zea mays* L.). *Genetics* 190, 795-811

Lau, N.C., Lim, L.P., Weinstein, E.G, Bartel, D.P. (2001) An Abundant class of Tiny RNAs with Probable Regulatory Roles in *Caenorhabditis elegans*. *Science* 294, 585-862

Law, J.A., Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* 11 (3), 204-220

Lee, R.C., Feinbaum, R.L., Ambros, V. (1993) The *C. elegans* Heterochronic Gene lin-4 Encodes Small RNAs with Antisense Complementarity to lin-14. *Cell* 75, 843-854

Lee, R.C., Ambros, V. (2001) An Extensive Class of Small RNAs in *Caenorhabditis elegans*. *Science* 294, 862-864

Li, J., Yang, Z., Yu, B., Chen, X. (2005) Methylation protects miRNAs and siRNAs from a 3'end uridylation activity in *Arabidopsis. Current Biology* 15 (16), 1501-1507

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25, 2078-2079

Li, Y., Varala, K., Moose, S.P., Hudson, M.E. (2012) The Inheritance Pattern of 24 nt siRNA Clusters in Arabidopsis Hybrids is Influenced by Proximity to Transposable Elements. *PloS ONE* 7, e47043

Li, L.Y., Yang, C., He, Y., Fang, R.Q., Tian, Z.H., Li, J.X. (2014) Expression patterns of microRNAs in different organs and developmental stages of a superhybrid rice LYP9 and its parental lines. *Plant Biology*, doi:10.1111/plb.12134

Lisch, D. (2009) Epigenetic Regulation of Transposable Elements in Plants. *Annual Review of Plant Biology* 60, 43-66

Liu, J., Han, L., Chen, F., Bao, J.; Zhang, F., Mi, G. (2008a) Microarray analysis reveals early responsive genes possibly involved in localized nitrate stimulation of lateral root development in maize (*Zea mays L.*). *Plant Science* 175 (3), 272-282

Liu, X., Fu, J., Gu, D., Liu, W., Liu, T., Peng, Y., Wang, J., Wang, G. (2008b) Genome-wide analysis of gene expression profiles during the kernel development of maize (*Zea mays* L.). *Genomics* 91 (4), 378-387

Liu, Y.X., Wang, M., Wang, X.J. (2014) Endogenous Small RNA Clusters in Plants. *Genomics, Proteomics & Bioinformatics* 12 (2), 64-71

Llave, C., Xie, Z., Kasschau, K.D., Carrington, J.C. (2002a) Cleavage of *Scarecrow-like* mRNA Targets Directed by a Class of Arabidopsis miRNA. *Science* 297, 2053-2056

Llave, C., Kasschau, K.D., Rector, M.A., Carrington, J.C. (2002b) Endogenous and Silencing-Associated Small RNAs in Plants. *The Plant Cell* 14, 1605-1619

Lu, C., Tej, S.S., Luo, S., Haudenschild, C.D., Meyers, B.C., Green, P.J. (2005) Elucidation of the Small RNA Component of the Transcriptome. *Science* 309, 1567-1569

Luo, M., Brown, R.L., Chen, Z.Y., Menkir, A., Yu, J., Bhatnagar, D. (2011) Transcriptional Profiles Uncover *Aspergillus flavus*-Induced Resistance in Maize Kernels. *Toxins* 3 (7), 766-786

Marí-Ordóñez, Marchais, A., Etcheverry, M., Martin, A., Colot, V., Voinnet, O. (2013) Reconstructing *de novo* silencing of an active retrotransposon. *Nature Genetics* 45, 1029-1039

Maron, L.G., Kirst, M., Mao, C., Milner, M.J., Menossi, M., Kochian, L.V. (2008) Transcriptional profiling of aluminium toxicity and tolerance responses in maize roots. *New Phytologist* 179 (1), 116-128

Martínez de Alba, A.E., Elvira-Matelot, E., Vaucheret, H. (2013) Gene silencing in plants: A diversity of pathways. *Biochimica et Biophysica Acta* 1829 (12), 1300-1308

Matzke, M., Mosher, R. (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics* 15 (6), 394-408

McClintock, B. (1984) The significance of response of the genome to challenge. *Science* 226, 792-801

McCormick, K.P., Willmann, M.R., Meyers, B.C. (2011) Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiments. *Silence* 2: 2

McCue, A.D., Nuthikattu, S., Reeder, S.H., Slotkin, R.K. (2012) Gene Expression and Stress Response Mediated by the Epigenetic Regulation of a Transposable Element Small RNA. *PloS Genetics* 8 (2), e1002474

McCue, A.D., Nuthikattu, S., Slotkin, R.K. (2013) Genome-wide identification of genes regulated in *trans* by transposable element small interfering RNAs. *RNA Biology* 10 (8), 1379-1395

McMullen, M.D., Kresovich, S., Sanchez Villeda, H., Bradbury, P., Li, H., Sun, Q., Flint-Garcia, S., Thornsberry, J., Acharya, C., Bottoms, C., Brown, P., Browne, C., Eller, M., Guill, K., Harjes, C., Kroon, D., Lepak, N., Mitchell, S.E., Peterson, B., Pressoir, G., Romero, S., Oropeza Rosas, M., Salvo, S., Yates, H., Hanson, M., Jones, E., Smith, S., Glaubitz, J.C., Goodman, M., Waren, D., Holland, J.B., Buckler, E.S. (2009) Genetic Properties of the Maize Nested Association Mapping Population. *Science* 325, 737-740

Medini, D., Donati, C., Tettelin, H., Masignani, V., Rappuoli, R. (2005) The microbial pangenome. *Current Opinions in Genetics & Development* 15 (6), 589-594

Melchinger, A.E., Gumber, R.K. (1998) Overview of heterosis and heterotic groups in agronomic crops. In: Lamkey, K.R, Staub, J.E. (eds.), Concepts and Breeding of Heterosis in Crop Plants, CSSA, Madison, WI, USA

Melchinger, A.E. (1999) Genetic Diversity and Heterosis. In: CIMMYT (1997) Book of Abstracts: The Genetics and Exploitation of Heterosis in Crops; An International Symposium. Mexico, D.F., Mexico

Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A., Matzke, A.J.M. (2000) Transcriptional silencing and promoter methylation triggered by double stranded RNA. *The EMBO Journal* 19 (19), 5194-5201

Meyer, S., Pospisil, H., Scholten, S. (2007) Heterosis associated gene expression in maize embryos 6 days after fertilization exhibits additive, dominant and overdominant pattern. *Plant Molecular Biology* 63, 381-391

Meyer, S. (2007) Heterosis assoziierte Genexpression in frühen Entwicklungsstadien von Zea mays L. Dissertation, Universität Hamburg: 133 pages

Michalak, P. (2009) Epigenetic, transposon and small RNA determinants of hybrid dysfunction. *Heredity* 102, 45-50

Mlotshwa, S., Pruss, G.J., Peragine, A., Endres, M.W., Li, J., Chen, X., Poethig, R.S., Bowman, L.H., Vance, V. (2008) DICER-LIKE2 Plays a Primary Role in Transitive Silencing of Transgenes in *Arabidopsis*. *PloS ONE* 3 (3), e1755

Moll, R.H., Lindsey, M.F., Robinson, H.F (1963) Estimates of Genetic Variances and Level of Dominance in Maize. *Genetics* 49, 411-423

Moll, R.H., Lonnquist, J.H., Vélez Fortuno, J., Johnson, E.C. (1965) The Relationship of Heterosis and Genetic Divergence in Maize. *Genetics* 52, 139-144

Moose, S., Boddu, J. (2010a) http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?GSE17932

Moose, S., Boddu, J. (2010b) http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?GSE17953

Moose, S., Church, J. (2009) http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?GSE18011

Moose, S., Zhao, H. (2009) http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?GSE18006

Moose, S., Zhao, H. (2010) http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?GSE18008

Moriles, J., Hansen, S., Horvath, D., Reicks, G., Clay, D., Clay, S. (2012) Microarray and Growth Analyses Identify Differences and Similarities of Early Corn Response to Weeds, Shade, and Nitrogen Stress. *Weed Science* 60 (2), 158-166

Morrison, K.M., Simmons, S.J., Stapleton, A.E. (2010) Loci controlling nitrate reductase activity in maize: ultraviolet-B signaling in aerial tissues increases nitrate reductase activity in leaf and root when responsive alleles are present. *Physiologia plantarum* 140 (4), 334-341

Mosher, R.A., Schwach, F., Studholme, D., Baulcombe, D.C. (2008) Pol IVb influences RNAdirected DNA methylation independently of its role in siRNA biogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 105 (8), 3145-3150

Mroczek, R.J., Dawe, R.K. (2003) Distribution of Retroelements in Centromeres and Neocentromeres of Maize. *Genetics* 165, 809-819

Murashige, T., Skoog, F. (1962) A Revised medium for Rapid Growth and Bio Assay with Tobacco Tissue. *Physiologia Plantarium* 15, 473-497

Napoli, C., Lemieux, C., Jorgensen, R. (1990) Introduction of a Chimeric Chalcone Synthase Gene into *Petunia* results in Reversible Co-Suppression of Homologous Genes in trans. *The Plant Cell* 2, 279-289

Nobuta, K., Lu, C., Shrivastava, R., Pillay, M., De Paoli, E., Accerbi, M., Arteaga-Vazquez, M., Sidorenko, L., Jeong, D.H., Yen, Y., Green, P.J., Chandler, V.L., Meyers, B.C. (2008) Distinct size distribution of endogenous siRNas in maize: Evidence from deep sequencing in the *mop1-1* mutant. *Proceedings of the National Academy of Sciences of the United States of America* 105 (39), 14958-14963

Nuthikattu, S., McCue, A.D., Panda, K., Fultz, D., DeFraia, C., Thomas, E.N., Slotkin, R.K. (2013) The Initiation of Epigenetic Silencing of Active Transposable Elements Is Triggered by RDR6 and 21-22 Nucleotide Small Interfering RNAs. *Plant physiology* 162 (1), 116-131

Ohtsu, K., Smith, M.B., Emrich, S.J., Borsuk, L.A., Zhou, R., Chen, T., Zhang, X., Timmermans, M.C.P., Beck, J., Buckner, B., Janick-Buckner, D., Nettleton, D., Scanlon, M.J., Schnable, P.S (2007) Global gene expression analysis of the shoot apical meristem of maize (*Zea mays* L.). *The Plant Journal* 52, 391-404

Park, W., Li, J., Song, R., Messing, J., Chen, X. (2002) CARPEL FACTORY, a Dicer Homolog, and HEN1, a novel Protein, Act in microRNA Metabolism in *Arabidopsis thaliana. Current Biology* 12, 1484-1495

Park, M.W., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., Poethig, R.S. (2005) Nuclear processing and export of microRNAs in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 102 (10), 3691-3696

Parvez, S. (2006) Recent advances in understanding genetic basis of heterosis in rice (*Oriza sativa* L.). *Revista UDO Agrícola* 6 (1), 1-10

Peláez, P., Sanchez, F. (2013) Small RNAs in plant defense responses during viral and bacterial interactions: similarities and differences. *Frontiers in Plant Science* 4: 343

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L., Poethig, R.S. (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of *trans*-acting siRNAs in *Arabidopsis. Genes & Development* 18, 2368-2379

Phimister, D. (1999) Going global. Nature Genetics 21 (supplement), 1

Pumplin, N., Voinnet, O. (2013) RNA silencing suppression by plant pathogens: defence, counter-defence and count-counter-defence. *Nature Reviews Microbiology* 11, 745-760

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Regulski, M., Lu, Zhenyuan, Kendall, J., Donoghue, M.T.A., Reinders, J., Llaca, V., Deschamps, S., Smith, A., Levy, D., McCombie, W.R., Tingey, S., Rafalski, A., Hicks, J., Ware, D., Martienssen, R.A. (2013) The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. *Genome Research* 23, 1651-1662

Reif, J.C., Melchinger, A.E., Xia, X.C., Warburton, M.L., Hoisington, D.A., Vasal, S.K., Srinivasan, G., Bohn, M., Frisch, M. (2003) Genetic Distance Based on Simple Sequence repeats and Heterosis in tropical Maize Populations. *Crop Science* 43, 1275-1282

Reif, J.C., Hallauer, A.R., Melchinger, A.E. (2005) Heterosis and heterotic patterns in maize. *Maydica* 50, 215-223

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., Bartel, D.P. (2002) microRNAs in plants. *Genes & Development* 16, 1616-1626

Riddle, N.C., Jiang, H., An, L., Doerge, R.W., Birchler, J.A. (2010) Gene expression analysis at the intersection of ploidy and hybridity in maize. *Theoretical and Applied Genetics* 120 (2), 341-355

Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., Sulpice, R., Altmann, T., Stitt, M., Willmitzer, L., Melchinger, A.E. (2013) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature Genetics* 44 (2), 217-220

Robinson, M.D., Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* 11: R25

Romagnoli, S., Maddaloni, M., Livini, C., Motto, M. (1990) Relationship between gene expression and hybrid vigor in primary root tips of young maize (*Zea mays* L.) plantlets. *Theoretical and Applied Genetics* 80, 769-775

Sawers, R.J.H., Liu, P., Anufrikova, K., Hwang, J.T.G., Brutnell, T.P. (2007) A multitreatment experimental system to examine photosynthetic differentiation in the maize leaf. *BMC Genomics* 8: 12
Schatz, M.C., Witkowski, J., McCombie, W.R. (2012) Current challenges in *de novo* plant genome sequencing and assembly. *Genome Biology* 13 (4): 243

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., Wilson, R.K. (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science 326, 1112-1115

Schnable, P.S., Springer, N.M. (2013) Progress Toward Understanding Heterosis in Crop Plants. *Annual review of plant biology* 64, 71-88

Schön, C.C., Dhillon, B.S., Utz, H.F., Melchinger, A.E. (2010) High congruency of QTL positions for heterosis of grain yield in three crosses of maize. *Theoretical and Applied Genetics* 120, 321-332

Schrag, T.A., Melchinger, A.E., Sørensen, A.P., Frisch, M. (2006) Prediction of single-cross hybrid performance for grain yield and grain dry matter content in maize using AFLP markers associated with QTL. *Theoretical and Applied Genetics* 113 (6), 1037-1047

Schrag, T.A., Maurer, H.P., Melchinger, A.E., Piepho, H.P., Peleman, J., Frisch, M. (2007) Prediction of single-cross hybrid performance in maize using haplotype blocks associated with QTL for grain yield. *Theoretical and Applied Genetics* 114, 1345-1355

Sentner, U. (2008) P1/HC-Pro aus dem Weizenstrichelmosaikvirus als Suppressoren des RNA *silencing* in Weizen. Dissertation, Universität Bayreuth: 150 pages

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

Shen, H., He, H., Li, J., Chen, W., Wang, X., Guo, L., Peng, Zhiyu, He, G., Zhong, S., Qi, Y., Terzaghi, W., Deng, X.W. (2012) Genome-Wide Analysis of DNA Methylation and Gene Expression Changes in Two *Arabidopsis* Ecotypes and Their Reciprocal Hybrids. *The Plant Cell* 24 (3), 875-892

Shull, G.H. (1908) The composition of a field of maize. *American Breeders Association Report*, 296–301

Shull, G.H. (1909) A pure line method of corn breeding. *American Breeders Association Report*, 51–59

Shull, G.H. (1914) Duplicate genes for capsule-form in *Bursa bursa-pastoris*. *Zeitschrift für Induktive Abstammungs- und Vererbungslehre* 12 (2), 97–149.

Slotkin, R.K., Martienssen, R.A. (2007) Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* 8, 272-285

Song, R., Messing, J. (2003) Gene expression of a gene family in maize based on noncollinear haplotypes. *Proceedings of the National Academy of Sciences of the United States of America* 100 (15), 9055-9060

Song, R., Qu, H., Chen, C., Hu, S.Y.J. (2007) Differential gene expression in an elite rice cultivar (*Oryza sativa*, L) and its parental lines based on SAGE data. *BMC Plant Biology* 7, 49.

Soós, V., Sebestyén, E., Juhász, A., Light, M.E., Kohout, L., Szalaj, G., Tandori, J., van Staden, J., Balász, E. (2010) Transcriptome analysis of germinating maize kernels exposed to smoke-water and the active compound KAR1. *BMC Plant Biology* 10: 236

Spollen, W.G., Tao, W., Valliyodan, B., Chen, K., Hejlek, L.G., Kim, J.J., LeNoble, M.E., Zhu, J. (2008) Spatial distribution of transcript changes in the maize primary root elongation zone at low water potential. *BMC Plant Biology* 8: 32

Springer, N.M., Stupar, R.M. (2007) Allelic variation and heterosis in maize: how do two halves make more than a whole? *Genome Research* 17 (3), 264-275

Springer, N.M., Ying, K., Fu, Y., Ji, T., Yeh, C.T., Jia, Y., Wu, W., Richmond, T., Kitzman, J., Rosenbaum, H., Iniguez, A., Barbazuk, W.B., Jeddeloh, J.A., Nettleton, D., Schnable, P.S. (2009) Maize Inbreds Exhibit High Levels of Copy Number Variation (CNV) and Presence/Absence Variation (PAV) in Genome Content. *PloS Genetics* 5 (11), e1000734

Stupar, R.M., Springer, N.M. (2006) Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. *Genetics* 173, 2199-2210

Stupar, R.M., Gardiner, J.M., Oldre, A.G., Haun, W.J., Chandler, V.L., Springer, N.M. (2008) Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. *BMC Plant Biology*8: 33

Swanson-Wagner, R.A., Jia, Y., DeCook, R., Borsuk, L.A., Nettleton, D., Schnable, P.S. (2006) All possible modes of gene action are observed in a global comparison of gene expression in a maize F_1 hybrid and its inbred parents. *Proceedings of the National Academy of Sciences of the United States of America* 103 (18), 6805-6810

Technow, F., Schrag, T.A., Schipprack, W., Bauer, E., Simianer, H., Melchinger, A.E. (2014) Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. *Genetics*, doi:10.1534/genetics.114.165860

Tenaillon, M.I., Sawkins, M.C., Long, A.D., Gaut, R.L., Doebley, J.F., Gaut, B.S. (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proceedings of the National Academy of Sciences of the United States of America* 98 (16), 9161-9166

Thiemann, A., Fu, J., Schrag, T.A., Melchinger, A.E., Frisch, M., Scholten, S. (2010) Correlation between parental transcriptome and field data for the characterization of heterosis in *Zea mays* L., *Theoretical and Applied Genetics* 120, 401-413

Thiemann, A. (2011) Genomweite Transkriptionsprofile zur Charakterisierung und Vorhersage von Heterosis in *Zea mays* L. Dissertation, Universität Hamburg: 123 pages

Thiemann, A., Fu, J., Seifert, F., Grant-Downton, R.T., Schrag, T.A., Pospisil, H., Frisch, M., Melchinger, A.E., Scholten, S. (2014) Genome-wide meta-analysis of maize heterosis re-

veals the potential role of additive gene expression at pericentromeric loci. *BMC Plant Biology* 14:88

Thieme, C.J., Schudoma, C., May, P., Walther, D. (2012) Give it AGO: The search for miRNA-Argonaute sorting signals in *Arabidopsis thaliana* indicates a relevance of sequence position other than the 5'-position alone. *Frontiers in Plant Science* 3, 272

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

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

Tsaftaris, S.A. (1995) Molecular aspects of heterosis in plants. *Physiologia Plantarum* 94, 362-370

Vaistij, F.E., Jones, L., Baulcombe, D.C. (2002) Spreading of RNA Targeting and DNA Methylation in RNA Silencing Requires Transcription of the Target Gene and a Putative RNA-Dependent RNA Polymerase. *The Plant Cell* 14, 857-867

Vance, V., Vaucheret, H. (2001) RNA Silencing in Plants - Defense and Counterdefense. Science 292, 2277-2280

van der Krol, A., Mur, L.A., Beld, m., Mol, J.N.M., Stuitje, A.R. (1990) Flavonoid Genes in *Petunia* - Addition of a Limited Number of Gene Copies May Lead to a Suppression of Gene Expression. *The Plant Cell* 2, 291-299

Várallyay, É., Válóczi, A., Ágyi, Á., Burgyán, J., Havelda, Z. (2010) Plant virus-mediated induction of miR168 is associated with repression of ARGONAUTE1 accumulation. *The EMBO Journal* 29, 3507-3519

Vasquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilber, J.L., Bartel, D.P., Créte, P. (2004) Endogenous trans-Acting siRNAs Regulate the Accumulation of Arabidopsis mRNAs. *Molecular Cell* 16, 69-79

Vasquez, F., Hohn, T. (2013) Biogenesis and Biological Activity of Secondary siRNAs in Plants. *Scientifica*, 783253

Vicient, C.M. (2010) Transcriptional activity of transposable elements in maize. *BMC Genomics* 11 (1): 601

Vogler, H., Akbergenov, R., Shivaprasad, P.V., Dang, V., Fasler, M., Kwon, M.O., Zhanybekova, S., Hohn, T., Heinlein, M. (2007) Modification of small RNAs associated with suppression of RNA silencing by tobamovirus replicase protein. *Journal of Virology* 81 (19), 10379-10388

Wei, L., Gu, L., Song, X., Cui, X., Lu, Z., Zhou, M., Wang, L., Hu, F., Zhai, J., Meyers, B.C., Cao, X. (2014) Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. *Proceedings of the National Academy of Sciences of the United States of America* 111 (10), 3877-3882

Wierzbicki, A.T., Haag, J.R., Pikaard, C.S. (2008) Noncoding transcription by RNA Polymerase Pol IVb/Pol V mediated transcriptional silencing of overlapping and adjacent genes. *Cell* 135 (4), 635-648

Wightman, B., Ha, I., Ruvkun, G. (1993) Posttranscriptional Regulation of the Heterochronic Gene lin-14 by lin-4 Mediates Temporal Pattern Formation in *C. elegans. Cell 75, 855-862*

Wu, L., Zhou, H., Zhang, Q., Zhang, J., Ni, F., Liu, C., Qi, Y. (2010) DNA Methylation Mediated by a MicroRNA Pathway. *Molecular Cell* 38 (3), 465-475

Wu, L., Mao, L., Qi, Y. (2012) Roles of Dicer-like and Argonaute proteins in TAS-derived small interfering RNA-triggered DNA methylation. *Plant Physiology* 160, 990-999

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., Carrington, J. (2004) Genetic and Functional Diversification of Small RNA Pathways in Plants. *PLoS Biology* 2 (5), 642-652

Xie, Z., Wang, J., Cao, M., Zhao, C., Zhao, K., Zshao, J., Lei, T., Xu, N., Liu, S. (2006) Pedigree analysis of an elite rice hybrid using proteomic approach. *Proteomics* 6 82), 474-486

Xie, Z., Qi, X. (2008) Diverse small RNA-directed silencing pathways in plants. Biochimica et Biophysica Acta 1779, 720-724

Xiong, L.Z., Xu, C.G., Saghai Maroof, M.A., Zhang, Q. (1999) Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique. *Molecular & General Genetics* 261, 439-446

Yang, Z., Ebright, Y.W., Yu., B., Chen, X. (2006) HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Research* 34 (2), 667-675

Yao, Y., Ni, Z., Zhang, Y., Chen, Y., Ding, Y., Han, Z., Liu, Z.L., Sun, Q. (2005) identification of differentially expressed genes in leaf and root between wheat hybrid and its parental inbreds using PCR-based cDNA subtraction. *Plant Molecular Biology* 58, 367-384

Yoshikawa, M., Iki, T., Tsutsui, Y., Miyashita, K., Poethig, R.S., Habu, Y., Ishikawa, M. (2013) 3' fragment of miR173-programmed RISC-cleaved RNA is protected from degradation in a complex with RISC and SGS3. *Proceedings of the National Academy of Sciences of the United States of America* 110 (10), 4117-4122

Young, B.A., Stenger, D.C., Qu, F., Morris, T.J., Tatieni, S., French, R. (2012) Tritimovirus P1 functions as a suppressor of RNA silencing and an enhancer of disease symptoms. *Virus Research* 163, 672-677

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., Chen, X. (2005) Methylation as a Crucial Step in Plant microRNA Biogenesis. *Science* 307, 932-935

Yue, G., Zhuang, Y., Li, Z., Sun, L., Zhang, J. (2008) Differential gene expression analysis of maize leaf at heading stage in response to water-deficit stress. *Bioscience Reports* 28 (3), 125-134

Zhang, L., Chia, J.M., Kumari, S., Stein, J.C., Liu, Z., Narechania, A., Maher, C.A., Guill, K., McMullen, M.D., Ware, D. (2009) A Genome-Wide Characterization of MicroRNA Genes in Maize. *PLoS Genetics* 5 (11), e1000716

Zhang, X., Xia, J., Lii, Y.E., Barrera-Figueroa, B.E., Zhou, X., Gao, S., Lu, L., Niu, D., Chen, Z., Leung, C., Wong, T., Zhang, H., Guo, J., Li, Y., Liu, R., Liang, W., Zhu, J.K., Zhang, W., Jin, H. (2012) Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. *Genome Biology* 13: R20

Zhang, S., Gardiner, J., Xiao, Y., Zhao, J., Wang, F., Zheng, Y. (2013) Floral transition in maize infected with *Sporisorium reilianum* disrupts compatibility with this biotrophic fungal pathogen. *Planta* 237 (5), 1251-1266

Zhaoxia, L., Changzheng, X., Kunpeng, L., Shi, Y., Xun, Q., Juren, Z. (2012) http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE36368 Zhou, S., Wei, F., Nguyen, J., Bechner, M., Potamousis, K., Goldstein, S., Pape, L., Mehan, M.R., Churas, C., Pasternak, S., Forrest, D.K., Wise, R., Ware, D., Wing, R.A., Waterman, M.S., Livny, M., Schwartz, D.C. (2009) A Single Molecule Scaffold for the Maize Genome. PloS Genetics 5 (11), e1000711

Zilberman, D., Cao, X., Jacobsen, S.E. (2003) ARGONAUTE4 Control of Locus-Specific siRNA Accumulation and DNA and Histone Methylation. *Science* 299, 716-719

X. Appendix

Appendix Table 1: Published experiments performed on either the 57K or 46K maizearray

GEO acc.	array	author	year	experimental focus
GSE3890	57K	Sawers et al.	2007	leaf development/photo- synthesis
GSE9341 GSE9352 GSE9369 GSE9379	57K	Spollen <i>et al.</i>	2008	root development/water stress
GSE9386	57K	Liu <i>et al.</i>	2008b	kernel development
GSE9453	57K	Fernandes et al.	2008	stress response
GSE9698	57K	Covshoff et al.	2008	photosynthesis
GSE10308	57K	Maron <i>et al.</i>	2008	aluminium stress
GSE10449	46K	Holding <i>et al.</i>	2008	opaque 2 modifier
GSE10542 GSE10543	46K	Stupar <i>et al.</i>	2008	heterosis
GSE10544	57K	Morrison et al.	2010	nitrate reductase regulation
GSE10596	57K	Yue et al.	2008	water stress
GSE11145	57K	Liu <i>et al.</i>	2008a	root development/nitrate
GSE14728	57K	Hayano- Kanashiro <i>et al.</i>	2009	water stress
GSE15853	57K	Barros <i>et al.</i>	2010	comparison GM/ non-GM
GSE17484	46K	Soós et al.	2010	germination/stress
GSE17754	46K	Thiemann <i>et al.</i>	2010	heterosis
GSE17932	57K	Moose & Boddu	2010*	developing earshoot
GSE17953	57K	Moose & Boddu	2010*	developing earshoot
GSE17971	57K	Moose & Zhao	2009*	leaf development
GSE18006	57K	Moose & Zhao	2010*	seed
GSE18008	57K	Moose & Ayodeji	2009*	developing earshoot
GSE18011	57K	Moose & Church	2009*	nitrogen response
GSE19883	46K	Luo <i>et al.</i>	2011	fungal resistance
GSE24014	46K	Bosch <i>et al.</i>	2011	cell wall biogenesis
GSE25526	57K	Riddle et al.	2010	ploidy and hybridity
GSE27709	46K	Johnson <i>et al.</i>	2011	corn earworm resistance
GSE29132	46K	Moriles <i>et al.</i>	2012	stress response
GSE33494	46K	Hansen <i>et al.</i>	2011*	water stress
GSE36368	46K	Zhaoxia <i>et al.</i>	2012*	root development/phosphate
GSE40107	46K	Johnson & Dowd	2012*	insect resistance
GSE40288	46K	Campos- Bermudez <i>et al.</i>	2012	fungal resistance
GSE41916	46K	Zhang et al.	2013	fungal resistance
GSE47982	46K	Engelberth	2013a*	stress defense
GSE50981	46K	Engelberth	2013b*	stress defense

* only dataset was published on NCBI GEO

Appendix Table 2: Hybrid trait values for MPH for GY [Mg ha-1 adjusted to 155 g kg-1 grain moisture] of 98 hybrids resulting from 14x7 inbred combinations

		paternal inbred parent						
		F037	F039	F043	F047	L024	L035	L043
	P033	57.645	52.739	55.18	55.125	52.684	53.009	48.2
	P040	49.499	49.537	50.263	45.883	46.38	46.471	49.923
	P046	59.263	53.888	55.791	56.187	54.091	55.04	51.641
	P048	51.051	51.411	52.233	48.841	49.2	50.31	52.805
rent	P063	57.536	52.045	55.564	54.657	53.191	54.253	49.191
d pa	P066	49.291	49.316	51.51	48.014	48.475	48.492	51.786
bred	S028	65.119	60.228	61.902	62.986	60.488	61.356	55.632
al in	S035	55.826	56.5	57.886	54.737	55.168	55.41	58.01
erna	S043	50.903	46.573	48.16	49.217	47.603	47.199	42.881
mat	S046	42.577	43.403	43.51	40.072	41.068	42.323	44.858
	S049	55.933	50.366	53.831	53.662	51.785	52.731	47.683
	S050	48.007	49.249	49.301	45.869	46.274	47.117	50.149
	S058	58.935	53.913	56.261	56.773	54.635	55.177	49.128
	S067	50.2	51.994	51.861	48.931	48.895	50.593	53.13

Appendix Table 3: Growth-rates of induced/non-induced transgenic P1/HC-Pro inbred and hybrid plants [cm/2 d]

indu	iced	non-induced			
inbred	hybrid	inbred	hybrid		
1.2	1.4615	1.65	2.3893		
1.5464	1.9821	2.115	0.575		
1.4786	1.1385	2.32	2.125		
1.1	1.9893	1.55	2.2036		
1.3596	2.2607	1.8393	1.762		
0.6892	1.4143	1.2322	1.7692		
1.306	1.6453	1.1972	1.5098		
1.4577	1.1646	1.1053	1.6157		
0.7143	1.5203	1.5967	1.3465		
1.4253	1.7882	1.8816	1.013		
1.7083	1.5143	1.5697	2.1		
0.8589	1.41	1.2702	1.7179		
1.45	1.9393	2.1	1.7036		
1.8519	1.5924	1.3821	1.7069		
0.8904	2.0457	1.2875	1.685		
1.5875	0.5538	1.6566	0.8614		
1.3625	2.76	1.4286	1.57		
0.6571	1.2215	1.341	1.665		
1.0029	1.4565	1.3178	2.2279		
1.3622	1.2487	1.1343	0.635		
1.3777	1.2152	1.64	2.1079		
1.6185	1.1453	1.422	1.8623		
1.9655	1.0253	1.5305	2.3071		
0.6794	1.0557	1.621			
0.2714	1.9	1.0229			
1.0923	2.2321	1.6367			
1.3798	1.0036	0.8657			
0.725	1.165	1.7915			
1.6186	1.405				
0.7	1.3214				

XXXIX

treatment	replicate	number of raw reads	number of processed reads	number of distinct processed reads
induced	1	14,355,957	5,841,571	1,164,701
induced	2	13,646,170	5,377,296	1,881,481
induced	3	18,706,239	6,949,410	2,076,452
non-induced	1	13,618,872	6,540,306	2,287,955
non-induced	2	23,296,800	12,399,595	3,247,013
non-induced	3	14,105,653	5,379,299	2,138,807

Appendix Table 4: Raw and processed read counts from P1/HC-Pro hybrid sequencing

Appendix Table 5: Raw and processed sequence counts from inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of Hohenheim

sample	run	number of raw reads	number of processed reads	number of distinct processed reads
B73	А	16,103,827	12,511,557	1,685,637
F037	А	22,653,720	19,038,867	1,396,788
F039	А	12,848,402	9,703,097	1,470,746
F043	С	21,939,051	14,488,783	4,072,399
F047	А	15,269,024	12,681,993	2,557,259
L024	А	24,442,183	19,794,581	2,970,112
L035	А	37,307,793	31,142,343	4,286,727
L043	А	32,582,844	24,682,433	3,600,852
P033	А	17,119,925	12,052,946	2,217,321
P040	А	18,267,239	14,513,727	2,794,270
P046	А	20,963,127	17,371,388	3,184,178
P048	А	13,285,411	11,123,942	2,276,086
P063	А	14,112,422	9,145,421	1,826,393
P066	А	12,953,371	11,184,753	2,423,948
S028	А	14,132,611	10,466,762	1,675,266
S036	А	18,225,524	12,813,503	1,605,377
S044	А	19,053,823	16,066,868	2,212,081
S046	А	14,581,363	12,956,834	1,340,702
S049	А	21,233,470	16,805,847	2,872,341
S050	А	21,610,221	11,880,077	1,537,121
S058	А	35,305,865	31,084,539	3,139,157
S067	А	18,969,844	12,511,557	1,685,637
P033xF047	В	35,266,505	23,111,082	4,262,626
S028xF039	В	13,702,409	8,953,310	2,805,689
S028xL024	D	28,534,979	17,360,183	3,541,357

run: A – Eurofins MWG GmbH, 15.07.2011; B – Eurofins MWG GmbH, 26.10.2011; C – Eurofins MWG GmbH, 10.01.2012; D – LGC Genomics GmbH, 07.12.2012

sample	total read counts	quantile normalized total read counts
B73	12,511,557	14,745,159
F037	19,038,867	14,462,918
F039	9,703,097	14,539,834
F043	14,488,783	15,704,066
F047	12,681,993	15,342,273
L024	19,794,581	15,508,388
L035	31,142,343	15,720,249
L043	24,682,433	15,647,481
P033	12,052,946	15,146,697
P040	14,513,727	15,448,539
P046	17,371,388	15,566,526
P048	11,123,942	15,184,307
P063	9,145,421	14,862,739
P066	11,184,753	15,272,952
S028	10,466,762	14,736,033
S036	12,813,503	14,671,735
S044	16,066,868	15,143,134
S046	16,805,847	15,477,101
S049	12,956,834	14,399,870
S050	31,084,539	15,555,721
S058	11,880,077	14,606,209
S067	12,071,379	15,366,044
P033xF047	8,953,310	15,543,107
S028xF039	23,111,082	15,719,285
S028xL024	17,360,183	15,637,962

Appendix Table 6: Raw and quantile normalized total read counts of the inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of Hohenheim

Appendix Table 7: Results from sRNA population complexity analysis for induced/non-induced transgenic P1/HC-Pro plants (bootstrap analysis with 1000 runs)

	average n distinc	umber of t reads		
bootstrap subset size [M]	induced transgenic plants	non-induced transgenic plants	sRNA complexity re- duction in induced transgenic plants [%]	p-value
0.1	44,434.92	56,338.09	21.13	0.015
0.2	78,384.35	102,485.65	23.52	0.017
0.5	164,489.15	222,303.4	26.01	0.02
1	284,132.21	391,295.07	27.39	0.022
1.5	386,915.39	537,489.89	28.01	0.023
2	478,066.67	667,632.83	28.39	0.023
2.5	560,144.29	785,144.47	28.66	0.024
3	634,715.89	892,219.28	28.86	0.024
3.5	702,898.1	990,357.91	29.03	0.024
4	765,552.38	1,080,759.63	29.17	0.025
3.5	823,308.53	1,164,368.53	29.29	0.025
5	876,741.98	1,241,884.32	29.40	0.026

Appendix Table 8: sRNA length distribution of normalized read counts for induced/non-induced P1/HC-Pro transgenic hybrids

					normalized	read counts		
trea	tment			induced			non-induced	
repl	icate		1	2	3	1	2	3
		18	78,455.26	60,381.25	70,614	55,029.39	75,514.75	69,875.01
		19	82,443.73	60,274.38	65,120.85	54,282.66	54,433.49	68,522.37
_		20	79,084.71	59,868.48	65,250.92	77,279.42	60,210.85	65,913.62
[n		21	89,368.32	85,630.29	88,869.61	97,928.27	81,463.21	85,454.36
gth		22	85,694.62	85,704.23	83,590.96	108,033.12	85,152.8	86,394.64
enç		23	76,200.48	66,316.19	64,312.8	68,346.16	71,254.92	67,558.8
IAI		24	121,018.23	192,888.9	177,094.36	248,761.65	266,806.11	213,247.88
SRN SRN		25	48,045	42,617.18	38,332.31	38,279.47	44,149.41	36,661.9
••		26	43,017.05	32,279.08	27,834.55	26,480.07	31,563.69	26,483.76
		27	40,369.99	26,703.37	23,584.69	21,320.54	25,812.52	19,808.49
		28	19,326.2	14,902.21	11,844.05	11,046.74	17,421.54	8,825.6

normalized read counts

Appendix Table 9: sRNA read mapping distribution to the B73 reference genome of the inbred line B73, 21 inbred lines and 3 hybrids from the breeding program of the University of Hohenheim

	fraction of distinct reads [%]					
			r	napping	g positior	າຣ
sample	mapped	unmapped	1	2-10	11-100	>100
B73	82.03	17.97	64.60	8.83	6.94	1.66
F037	44.85	55.15	29.58	6.67	7.54	1.06
F039	47.85	52.15	31.46	7.14	8.04	1.21
F043	56.14	43.86	37.84	9.29	7.30	1.72
F047	53.53	46.47	36.83	8.05	7.31	1.33
L024	54.47	45.53	39.25	7.39	6.59	1.24
L035	50.14	49.86	35.71	7.21	5.90	1.32
L043	50.13	49.87	35.61	7.19	6.04	1.29
P033	57.48	42.52	40.39	8.21	7.42	1.47
P040	58.10	41.90	41.19	8.18	7.21	1.52
P046	57.85	42.15	42.06	7.95	6.54	1.30
P048	55.62	44.38	39.31	7.74	7.14	1.44
P063	52.07	47.93	36.56	7.08	7.15	1.28
P066	53.53	46.47	38.30	7.19	6.82	1.21
S028	55.55	44.45	37.51	8.15	8.54	1.35
S036	52.23	47.77	35.85	7.62	7.60	1.16
S044	51.28	48.72	35.59	7.22	7.10	1.38
S046	52.09	47.91	35.30	7.30	8.27	1.23
S049	53.33	46.67	37.95	7.35	6.72	1.31
S050	53.55	46.45	35.88	7.86	8.45	1.36
S058	52.59	47.41	38.53	6.69	6.19	1.18
S067	47.97	52.03	33.07	6.51	7.23	1.18
P033xf047	59.87	40.13	41.41	9.52	7.20	1.74
S028xf039	65.28	34.72	44.65	10.68	8.04	1.90
S028xI024	63.72	36.28	44.15	9.83	7.98	1.76

Appendix Table 10: Probabilities of differences in read mapping position counts between Flint and Dent lines as well as inbreds and hybrids

			mapping positions		
sample set comparison	mapped/ unmapped	1	2-10	11-100	>100
Flint/Dent	0.130	0.123	0.869	0.346	0.983
inbred/hybrid	0.011	0.005	0.009	0.165	0.001

P033

P040

P046

P049

P063

P066

S028

S036

S044

S046

S049

S050

S058

S067

Flint

both

Dent

Flint

both

Dent

Flint

both

Dent

Flint

both

Dent

31.69

35.13

33.19

34.36

34.51

31.12

31.33

34.81

33.86

34.31

34.86

30.83

29.92

33.31

36.77

33.54

33.54

32.92

29.87

33.28

36.86

33.26

33.67

33.07

33.15

31.95

34.9

37.44

33.09

29.47

32.64

31.46

37.49

33.55

28.97

35.9

dix Ta	dix Table 11: sRNA population overlaps between inbred parents [%]								
	F037	F039	F043	F047	L024	L035	L043		
Flint	33.34	32.5	33.31	30.14	30.93	31.87	30.79		
both	35.34	34.37	30.66	29.29	29.44	29.16	28.56		
Dent	31.32	33.13	36.02	40.57	39.63	38.97	40.65		
Flint	33.44	31.63	32.38	28.28	29.64	31.23	29.87		
both	36.79	34.93	31.17	28.94	29.64	30	29.14		
Dent	29.76	33.45	36.45	42.78	40.72	38.77	40.99		
Flint	33.21	32.08	32.33	28.84	29.9	30.42	29.94		
both	36.6	35.38	31.14	29.51	29.93	29.21	29.24		
Dent	30.2	32.55	36.53	41.65	40.17	40.37	40.82		
Flint	33.59	32.26	34.04	30.43	31.08	31.79	30.72		
both	35.58	34.14	31.42	29.58	29.59	29.08	28.49		
Dent	30.84	33.6	34.54	39.98	39.33	39.13	40.79		
Flint	33.92	33.52	35.4	32.84	33.74	33.37	32.59		
both	35.13	34.58	32.01	31.2	31.49	29.88	29.58		
Dent	30.95	31.89	32.59	35.96	34.78	36.75	37.83		
Flint	33.17	30.95	33.65	28.64	30.02	31.87	29.83		
both	35.81	33.54	31.7	28.49	29.23	29.86	28.29		
Dent	31.02	35.51	34.65	42.87	40.75	38.27	41.89		
Flint	34.4	32.97	34.58	30.57	32.51	33.91	32.13		
both	34.94	33.37	30.46	28.15	29.52	29.75	28.38		
Dent	30.67	33.66	34.96	41.28	37.97	36.34	39.49		
Flint	34.9	33.29	36.74	33.28	33.98	35.85	33.66		
both	35.19	33.45	32.5	30.69	30.79	31.55	29.73		
Dent	29.91	33.26	30.75	36.03	35.23	32.6	36.62		
Flint	32.94	32.6	35.97	32.95	33.46	33.34	32.16		
both	34.94	34.47	33.42	32.13	32.03	30.69	29.97		
Dent	32.12	32.93	30.61	34.93	34.51	35.97	37.88		
Flint	36.22	35.13	39.31	35.89	36.2	37.64	35.82		
both	36.12	34.88	34.84	33	32.71	33.04	31.6		
Dent	27.67	29.99	25.85	31.11	31.09	29.32	32.58		

28.29

28.95

42.76

34.09

31.39

34.52

27.43

28.12

44.45

34.61

32.33

33.06

29.54

29.54

40.93

34.81

33.69

30.01

30.04

39.95

35.09

32.21

32.69

31.5

31.06

29.83

39.11

35.7

31.24

33.06

30.82

29.61

39.56

31.97

32.03

36

28.34

27.59

44.07

33.95

29.88

36.17

29.11

28.39

42.5

33.8

30.14

36.06

Appendix

			distinct	reads	enrichment		
		ha-sR	NAs	bootstrap sample	proba	bility	
		positive	negative	(both pos./neg.)	positive	negative	
	18	54	83	458.43	1	1	
	19	48	93	484.34	1	1	
	20	66	102	535.02	1	1	
Ē	21	201	365	797.37	1	1	
gth	22	576	1,452	973.54	0	0	
enç	23	242	447	1,497.39	1	1	
AI	24	2,830	3,888	4,757.12	0	0	
RN N	25	117	133	894.78	1	1	
S	26	70	91	381.85	1	1	
	27	78	123	269.63	0.99	1	
	28	75	138	222.53	0.89	0.41	

Appendix Table 12: ha-sRNA length distribution and enrichment probabilities

Appendix Table 13: sRNA length distribution of pos. ha-sRNA for mapping position count groups on the B73 reference genome

	mapping position count on B73 reference genome									
sRNA length [nt]	unmapped	1	2-10	11-100	>100					
18	28	5	5	5	11					
19	20	7	7	4	10					
20	29	7	9	3	18					
21	74	22	38	21	46					
22	146	66	106	87	171					
23	98	36	43	27	38					
24	1,457	670	472	140	91					
25	43	13	23	12	26					
26	33	5	4	5	23					
27	43	5	1	2	27					
28	30	8	10	2	25					

XLIV

Appendix Table 14: sRNA length distribution of neg. ha-sRNAs for mapping position count groups on the B73 reference genome

	mapping posi	ition cou	unt on B73	3 reference	genome
sRNA length [nt]	unmapped	1	2-10	11-100	>100
18	20	8	19	11	25
19	26	9	8	6	44
20	18	10	19	9	46
21	70	41	41	49	164
22	162	108	197	201	784
23	174	76	73	48	76
24	2,003	901	592	229	163
25	47	11	23	10	42
26	24	8	10	11	38
27	44	9	16	16	38
28	47	5	24	12	50

Appendix Table 15: sRNA length distribution of pos. ha-sRNAs to annotation types

sRNA length [nt]	gene	repeat	interge nic	gene/ repeat	gene/ inter- genic	repeat/ inter- genic	gene/ repeat/ intergenic
18	0	2	6	0	6	12	0
19	3	1	10	0	3	11	0
20	0	3	13	0	2	19	0
21	9	10	30	2	21	28	27
22	19	34	105	9	59	57	147
23	3	16	55	3	14	29	23
24	46	269	682	32	93	148	98
25	2	2	26	2	7	27	8
26	0	4	8	0	1	24	0
27	0	2	5	0	1	27	0
28	1	2	13	0	1	28	0

sRNA length [nt]	gene	repeat	intergenic	gene/ repeat	gene/ inter- genic	repeat/ inter- genic	gene/ repeat/ intergenic
18	3	0	17	0	7	31	4
19	0	4	12	0	3	46	2
20	5	1	21	0	5	39	13
21	17	16	44	0	44	51	123
22	23	74	172	8	235	129	648
23	4	47	89	8	26	56	43
24	58	357	915	47	140	175	189
25	0	12	19	3	5	40	7
26	3	1	15	1	4	42	1
27	3	3	27	0	4	42	0
28	1	1	29	0	5	55	0

Appendix Table 16: sRNA length distribution of neg. ha-sRNAs to annotation types

Appendix Table 17: sRNA length distribution of ha-sRNAs to repeat super-families

sRNA length [nt]	LINE L1	LINE Unknown	LTR Copia	LTR Gypsy	LTR Unknown	TIR CACTA	тік һАТ	TIR Mutator	TIR Pif-Harbinger	TIR Tc1-Mariner
18	0	0	1	2	46	0	0	0	0	0
19	0	0	1	1	62	0	0	1	0	0
20	0	0	0	11	65	0	0	0	0	0
21	0	1	48	141	95	12	5	8	6	1
22	0	2	263	804	231	28	3	21	17	0
23	0	0	35	64	107	28	15	11	10	1
24	2	18	196	239	416	246	191	115	119	10
25	0	0	5	8	80	5	3	4	5	2
26	0	0	2	2	71	0	0	0	1	0
27	0	0	0	0	74	0	0	0	0	0
28	0	0	1	2	85	0	0	0	0	0

Appendix Table 18: Average sRNA length distribution to repeat super-families calculated from 1000 bootstrap runs

sRNA _length [nt]	LINE L1	LINE Unknown	LTR Copia	LTR Gypsy	LTR Unknown	TIR CACTA	TIR hAT	TIR Mutator	TIR Pif-Harbinger	TIR Tc1-Mariner
18	0.07	0.66	12.32	15.98	18.01	6.84	4.78	2.74	3.46	0.47
19	0.06	0.89	16.35	20.62	23.11	8.88	6.15	3.64	4.43	0.61
20	0.09	1.03	19.12	24.56	26.19	9.95	7.21	4.18	4.99	0.71
21	0.37	3.41	68.74	88.16	81.56	32.26	23.19	14.47	15.99	2.04
22	1.27	14.65	331.67	439.38	332.85	120.63	77.57	52.33	55.08	7.31
23	0.39	3.54	50.38	68.14	70.09	29.7	23.17	12.56	16.76	2.49
24	2.54	22.87	285.54	424.53	387.23	155.48	125.8	67.41	87.94	13.08
25	0.16	1.44	22.44	30.05	33.06	13.23	9.57	5.23	7.14	1.2
26	0.08	0.87	15.23	20.09	32.48	7.68	5.27	3.08	4.17	0.76
27	0.06	0.71	12.4	16	45.71	4.89	2.78	2.16	2.51	0.55
28	0.08	0.83	11.24	17.05	59.02	4.12	2.13	2.13	2.18	0.48

sRNA length [nt]

18 19

20 21

22 23

24

25

26

27

28

1/0.92

1/0.94

1/0.92

1/0.42

1/0.49

1/0.44

1/0

1/0

1/0

1/0

1/0

1/0

0/1

0/1

0/1

1/0

1/0

1/0.01

LINE L1	LINE Unknown	LTR Copia	LTR Gypsy	LTR Unknown	TIR CACTA	TIR hAT	TIR Mutator	TIR Pif/Harbinger	TIR Tc1/Mariner
1/0.94	1/0.5	1/0	1/0	0/1	1/0	1/0	1/0.06	1/0.03	1/0.62
1/0.94	1/0.42	1/0	1/0	0/1	1/0	1/0	0.98/ 0.12	1/0.01	1/0.54
1/0.92	1/0.35	1/0	1/0	0/1	1/0	1/0	1/0.02	1/0	1/0.49
1/0.71	0.97/ 0.14	1/0	0/1	0.04/ 0.98	1/0	1/0	0.98/ 0.04	1/0	0.89/ 0.4
1/0.27	1/0	1/0	0/1	1/0	1/0	1/0	1/0	1/0	1/0
1/0.68	1/0.03	0.99 /0.01	0.75/ 0.29	0/1	0.68/ 0.41	0.98/ 0.03	0.71/ 0.39	0.97/ 0.06	0.91/ 0.29
0.72/ 0.54	0.88/ 0.18	1/0	1/0	0.04/ 0.96	0/1	0/1	0/1	0/1	0.85/ 0.23
1/0.85	1/0.23	1/0	1/0	0/1	1/0.01	1/0.01	0.77/	0.86	0.33/

0.39

1/0.04

1/0.11

1/0.12

1/0

1/0.06

1/0.12

/0.26

0.99/

80.0

1/0.07

1/0.11

0.86

1/0.46

1/0.57

1/0.62

Appendix Table 19: sRNA length enrichment/depletion probabilities for ha-sRNA to repeat super-families

sRNA length [nt]	LINE L1	LINE Unknown	LTR Copia	LTR Gypsy	LTR Unknown	TIR CACTA	TIR hAT	TIR Mutator	TIR Pif/Harbinger	TIR Tc1/Mariner
18	0	0	0	0	14	0	0	0	0	0
19	0	0	0	0	12	0	0	0	0	0
20	0	0	0	0	22	0	0	0	0	0
21	0	0	15	22	31	11	2	4	0	0
22	0	0	77	131	74	20	1	16	5	0
23	0	0	19	18	30	12	3	6	2	0
24	1	7	87	97	172	104	89	44	47	8
25	0	0	2	5	31	3	0	1	1	1
26	0	0	0	0	27	0	0	0	1	0
27	0	0	0	0	29	0	0	0	0	0
28	0	0	0	0	30	0	0	0	0	0

Appendix Table 20: sRNA length distribution of pos. ha-sRNAs to repeat super-families

Appendix Table 21: sRNA length distribution of neg. ha-sRNAs to repeat super-families

sRNA length [nt]	LINE L1	LINE Unknown	LTR Copia	LTR Gypsy	LTR Unknown	TIR CACTA	TIR hAT	TIR Mutator	TIR Pif/Harbinger	TIR Tc1/Mariner
18	0	0	1	2	32	0	0	0	0	0
19	0	0	1	1	50	0	0	1	0	0
20	0	0	0	11	43	0	0	0	0	0
21	0	1	33	119	64	1	3	4	6	1
22	0	2	186	673	157	8	2	5	12	0
23	0	0	16	46	77	16	12	5	8	1
24	1	11	109	142	244	142	102	71	72	2
25	0	0	3	3	49	2	3	3	4	1
26	0	0	2	2	44	0	0	0	0	0
27	0	0	0	0	45	0	0	0	0	0
28	0	0	1	2	55	0	0	0	0	0

sRNA length [nt]	LINE L1	LINE Unknown	LTR Copia	LTR Gypsy	LTR Unknown	TIR CACTA	TIR hAT	TIR Mutator	TIR Pif-Harbinger	TIR Tc1-Mariner
18	1/ 0.98	1/ 0.82	1/ 0.01	1/0	0/1	1/ 0.14	1/ 0.25	1/ 0.41	1/ 0.39	1/ 0.88
19	1/ 0.99	1/ 0.87	1/ 0.03	1/0	0/1	1/ 0.15	1/ 0.31	1/0.5	1/ 0.41	1/ 0.89
20	1/ 0.97	1/ 0.76	1/0	1/0	0/1	1/ 0.05	1/ 0.09	1/ 0.27	1/ 0.22	1/ 0.81
21	1/ 0.91	1/ 0.4	0.84/ 0.25	0.66/ 0.46	0.01/ 0.99	0.17/ 0.91	0.99/ 0.04	0.51/ 0.66	1/ 0.01	1/ 0.58
22	1/ 0.76	1/ 0.05	0.35/ 0.69	0/1	0.56/ 0.5	0.95/ 0.1	1/ 0	0.13/ 0.92	1/ 0.01	1/0.2
23	1/ 0.9	1/ 0.35	0.22/ 0.85	0.87/ 0.2	0.03/ 0.98	0.23/ 0.86	0.98/ 0.06	0.2/ 0.9	0.98/ 0.07	1/ 0.47
24	0.64/ 0.71	0.84/ 0.27	1/0	1/0	0.17/ 0.85	0/ 1	0/ 1	0/1	0.06/ 0.96	0.19/ 0.91
25	1/ 0.95	1/ 0.59	1/0	0.99/ 0.02	0/ 1	0.9/ 0.24	1/ 0.03	0.88/ 0.41	0.94/ 0.21	0.37/ 0.93
26	1/ 0.97	1/ 0.72	1/0	1/0	0/ 1	1/ 0.04	1/ 0.13	1/0.3	0.79/ 0.51	1/ 0.74
27	1/ 0.98	1/ 0.72	1/ 0.01	1/0	0/ 1	1/ 0.17	1/ 0.34	1/ 0.41	1/ 0.35	1/ 0.82
28	1/ 0.97	1/ 0.77	1/ 0.02	1/0	0/ 1	1/ 0.21	1/ 0.45	1/ 0.43	1/ 0.48	1/ 0.84

Appendix Table 22: sRNA length enrichment/depletion probabilities for pos. ha-sRNA to repeat super-families

sRNA length [nt]	LINE L1	LINE Unknown	LTR Copia	LTR Gypsy	LTR Unknown	TIR CACTA	TIR hAT	TIR Mutator	TIR Pif-Harbinger	TIR Tc1-Mariner
18	1/ 0,98	1/ 0,82	1/ 0,01	1/0	0/1	1/ 0,14	1/ 0,25	1/ 0,41	1/ 0,39	1/ 0,88
19	1/ 0,99	1/ 0,87	1/ 0,03	1/0	0/1	1/ 0,15	1/ 0,31	1/0,5	1/ 0,41	1/ 0,89
20	1/ 0,97	1/ 0,76	1/0	1/0	0/1	1/ 0,05	1/ 0,09	1/ 0,27	1/ 0,22	1/ 0,81
21	1/ 0,91	1/ 0,4	0,84/ 0,25	0,66/ 0,46	0,01/ 0,99	0,17/ 0,91	0,99/ 0,04	0,51/ 0,66	1/ 0,01	1/ 0,58
22	1/ 0,76	1/ 0,05	0,35/ 0,69	0/1	0,56/ 0,5	0,95/ 0,1	1/0	0,13/ 0,92	1/ 0,01	1/0,2
23	1/ 0,9	1/ 0,35	0,22/ 0,85	0,87/ 0,2	0,03/ 0,98	0,23/ 0,86	0,98/ 0,06	0,2/ 0,9	0,98/ 0,07	1/ 0,47
24	0,64/ 0,71	0,84/ 0,27	1/0	1/0	0,17/ 0,85	0/1	0/1	0/1	0,06/ 0,96	0,19/ 0,91
25	1/ 0,95	1/ 0,59	1/0	0,99/ 0,02	0/1	0,9/ 0,24	1/ 0,03	0,88/ 0,41	0,94/ 0,21	0,37/ 0,93
26	1/ 0,97	1/ 0,72	1/0	1/0	0/1	1/ 0,04	1/ 0,13	1/0,3	0,79/ 0,51	1/ 0,74
27	1/ 0,98	1/ 0,72	1/ 0,01	1/0	0/1	1/ 0,17	1/ 0,34	1/ 0,41	1/ 0,35	1/ 0,82
28	1/ 0,97	1/ 0,77	1/ 0,02	1/0	0/1	1/ 0,21	1/ 0,45	1/ 0,43	1/ 0,48	1/ 0,84

Appendix Table 23: sRNA length enrichment/depletion probabilities for neg. ha-sRNA to repeat super-families

sRNA length [nt]	RLC Ji	RLC leviathan	RLC opie	RLG cinful-zeon	RLG doke	RLG gyma	RLG huck	RLG nihep	RLG prem1	RLG xilon-diguus	RLX gunu	RLX ipiki	RLX milt	RLX yraj
18	0	0	1	1	0	0	0	1	0	0	0	44	0	0
19	1	0	0	1	0	0	0	1	0	0	0	60	0	0
20	0	0	0	10	0	2	0	5	0	0	1	61	0	0
21	16	8	17	98	11	9	12	43	2	9	9	62	1	1
22	87	71	69	560	37	80	34	250	10	36	42	61	1	2
23	16	2	8	35	3	11	9	10	3	2	5	57	5	14
24	53	9	42	72	13	26	64	24	30	15	21	51	53	65
25	2	0	0	4	0	0	1	1	3	0	1	66	1	2
26	1	0	0	0	0	0	0	0	2	0	0	68	0	0
27	0	0	0	0	0	0	0	0	0	0	1	73	0	0
28	1	0	0	0	0	0	1	0	1	0	0	84	0	0

Appendix Table 24: sRNA length distribution of ha-sRNAs to repeat families

sRNA length [nt]	RLC ji	RLC leviathan	RLC opie	RLG cinful-zeon	RLG doke	RLG gyma	RLG huck	RLG nihep	RLG prem1	RLG xilon-diguus	RLX gunu	RLX ipiki	RLX milt	RLX yraj
18	1/ 0,014	1/ 0,449	0,977/ 0,109	0,986/ 0,071	1/ 0,211	1/ 0,169	1/ 0,01	0,815/ 0,463	1/ 0,265	1/ 0,281	1/ 0,076	0/1	1/ 0,194	1/ 0,517
19	1/ 0,015	1/ 0,346	1/ 0,01	0,991/ 0,025	1/ 0,128	1/ 0,114	1/ 0,005	0,917/ 0,29	1/ 0,215	1/ 0,205	1/ 0,037	0/1	1/ 0,133	1/ 0,409
20	1/0	1/ 0,324	1/ 0,005	0,106/ 0,941	1/ 0,071	0,697/ 0,555	1/ 0,004	0,126/ 0,953	1/ 0,16	1/ 0,172	0,981/ 0,105	0/1	1/ 0,094	1/ 0,386
21	0,971/ 0,044	0,058/ 0,977	0,656/ 0,438	0/1	0,411/ 0,707	0,493/ 0,632	0,97/ 0,047	0/1	0,973/ 0,081	0,174/ 0,897	0,905/ 0,172	0/1	1/ 0,002	0,921/ 0,248
22	1/0	0/1	0,98/ 0,026	0/1	0,981/ 0,027	0/1	1/0	0/1	1/ 0,001	0,128/ 0,901	0,999/ 0,001	0/1	1/0	1/ 0,001
23	0,421/ 0,673	0,767/ 0,485	0,961/ 0,067	0/ 1	0,87/ 0,28	0,057/ 0,979	0,973/ 0,058	0,047/ 0,981	0,921/ 0,172	0,958/ 0,111	0,945/ 0,118	0/1	0,903/ 0,181	0/1
24	0,999/ 0,001	0,947/ 0,1	1/0	0,997/ 0,005	0,999/ 0,002	0,988/ 0,024	0,996/ 0,008	0,592/ 0,479	0,942/ 0,082	1/ 0,001	1/0	0/1	0,231/ 0,807	0/1
25	0,994/ 0,031	1/ 0,325	1/ 0,003	0,913/ 0,191	1/ 0,156	1/ 0,074	0,999/ 0,007	0,903/ 0,353	0,473/ 0,759	1/ 0,111	0,982/ 0,092	0/1	0,976/ 0,114	0,399/ 0,851
26	0,99/ 0,052	1/ 0,447	1/ 0,01	1/ 0,011	1/ 0,344	1/ 0,094	1/ 0,002	1/ 0,221	0,456/ 0,797	1/ 0,176	1/ 0,049	0/1	1/ 0,163	1/ 0,396
27	1/ 0,016	1/ 0,53	1/ 0,03	1/ 0,055	1/ 0,418	1/ 0,085	1/ 0,021	1/ 0,347	1/ 0,217	1/ 0,239	0,918/ 0,311	0/1	1/ 0,362	1/ 0,577
28	0,964/ 0,164	1/ 0,525	1/ 0,051	1/ 0,056	1/ 0,432	1/ 0,031	0,981/ 0,107	1/ 0,314	0,836/ 0,472	1/ 0,238	1/ 0,088	0/1	1/ 0,391	1/ 0,57

Appendix Table 25: sRNA length enrichment/depletion probabilities for ha-sRNAs to repeat families

Appendix Table 26: sRNA length distribution of pos. ha-sRNAs to repeat families

sRNA length [nt]	RLC ji	בפסח zeon	RLX ipiki	TIR Mutator Zm00118	TIR Mutator Zm00754
18	0	0	14	0	0
19	0	0	10	0	0
20	0	0	22	0	0
21	9	5	19	0	0
22	50	28	22	0	0
23	9	11	14	13	52
24	22	20	17	63	0
25	0	2	22	0	0
26	0	0	26	0	0
27	0	0	29	0	0
28	0	0	30	0	0

Appendix Table 27: sRNA length enrichment/depletion probabilities pos. ha-sRNA to repeat families

sRNA length	ILC JI	RLG cinful- zeon	RLX ipiki	TIR Mutator Zm00118	TIR Mutator Zm00754
18	1/0.279	1/0.298	0/1	1/0.981	1/0.995
19	1/0.342	1/0.352	0/1	1/0.988	1/0.994
20	1/0.127	1/0.137	0/1	1/0.974	1/0.986
21	0.194/ 0.913	0.748/ 0.414	0/1	1/0.948	1/0.974
22	0/1	0.455/ 0.633	0/1	1/0.827	1/0.938
23	0.057/ 0.976	0.016/ 0.994	0/1	0.093/ 0.992	0.024/ 1
24	0.987/ 0.018	1/0	0/1	0/1	1/0.81
25	1/0.077	0.732/ 0.517	0/1	1/0.95	1/0.98
26	1/0.176	1/0.178	0/1	1/0.963	1/0.988
27	1/0.236	1/0.32	0/1	1/0.973	1/0.993
28	1/0.318	1/0.346	0/1	1/0.976	1/0.994

sRNA length [nt]	RLC ji	RLC leviathan	RLC opie	RLG cinful-zeon	RLG gyma	RLG huck	RLG nihep	RLG xilon-diguus	RLX ipiki	RLX yraj
18	0	0	1	1	0	0	1	0	30	0
19	1	0	0	1	0	0	1	0	50	0
20	0	0	0	10	2	0	5	0	39	0
21	7	8	12	93	9	9	40	8	43	1
22	37	67	40	532	69	24	233	34	39	2
23	7	2	4	24	10	7	9	2	43	11
24	31	6	26	52	17	32	19	9	34	37
25	2	0	0	2	0	0	0	0	44	1
26	1	0	0	0	0	0	0	0	42	0
27	0	0	0	0	0	0	0	0	44	0
28	1	0	0	0	0	1	0	0	54	0

Appendix Table 28: sRNA length distribution of neg. ha-sRNAs to repeat families

sRNA length [nt]	RLC ji	RLC leviathan	RLC opie	RLG cinful-zeon	RLG gyma	RLG huck	RLG nihep	RLG xilon-diguus	RLX ipiki	RLX yraj
18	1/ 0,033	1/ 0,586	0,933/ 0,248	0,952/ 0,18	1/ 0,288	1/ 0,053	0,705/ 0,651	1/0,4	0/1	1/ 0,582
19	0,993/ 0,037	1/ 0,453	1/ 0,024	0,989/ 0,051	1/ 0,186	1/ 0,017	0,852/ 0,437	1/0,28	0/1	1/ 0,488
20	1/ 0,003	1/ 0,439	1/0,02	0,017/ 0,992	0,522/ 0,752	1/ 0,009	0,048/ 0,989	1/ 0,256	0/1	1/ 0,507
21	1/0,00 2	0,006/ 0,997	0,689/ 0,413	0/1	0,187/ 0,889	0,937/ 0,11	0/1	0,091/ 0,958	0/1	0,859/ 0,423
22	1/0	0/1	1/0	0/1	0/1	1/0	0/1	0,012/ 0,993	0/1	0,992/ 0,029
23	0,893/ 0,182	0,524/ 0,699	0,995/ 0,028	0/1	0,014/ 0,994	0,895/ 0,201	0,012/ 0,997	0,865/ 0,313	0/1	0/1
24	1/ 0,003	0,835/ 0,291	1/ 0,001	0,753/ 0,301	0,884/ 0,168	1/0	0,151/ 0,9	0,991/ 0,017	0/1	0/1
25	0,915/ 0,218	1/ 0,514	1/ 0,018	0,925/ 0,201	1/0,17 6	1/0,01 5	1/0,25 9	1/0,25 8	0/1	0,585/ 0,786
26	0,945/ 0,204	1/ 0,588	1/ 0,063	1/0,062	1/0,24 5	1/0,02 5	1/0,35 3	1/0,38 7	0/1	1/0,55 3
27	1/ 0,121	1/ 0,664	1/0,14 8	1/0,162	1/0,18 5	1/0,11 4	1/0,55 6	1/0,39 8	0/1	1/0,71 5
28	0,887/ 0,367	1/ 0,649	1/0,13 6	1/0,156	1/ 0,119	0,915/ 0,301	1/ 0,438	1/ 0,393	0/1	1/ 0,711

Appendix Table 29: sRNA length enrichment/depletion probabilities neg. ha-sRNA to repeat families

sRNA length [nt]	all ha-sRNAs	pos. ha-sRNAs	neg. ha-sRNAs
18	0	0	0
19	0	0	0
20	0	0	0
21	6	3	3
22	29	15	15
23	0	0	0
24	3	3	0
25	0	0	0
26	0	0	0
27	0	0	0
28	0	0	0

Appendix Table 30: Length distribution of ha-sRNAs with expression significantly correlated to putative target transcripts

Appendix Table 31: sRNA length distribution of ha-sRNAs, present in the inbred lines B73 and Mo17

sRNA length [nt]	all ha-sRNAs	differentially expressed ha-sRNAs
18	81	6
19	98	3
20	106	2
21	299	7
22	1,327	31
23	277	10
24	2,111	24
25	129	0
26	107	4
27	134	10
28	139	12

XI. Declaration on oath / Eidesstattliche Versicherung

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hamburg, date / Hamburg, den

Signature / Unterschrift

XII. Confirmation of correct English

ST PETER'S COLLEGE Tel: 01865 278900 New Inn Hall Street Fax: 01865 278855 Oxford www.spc.ox.ac.uk OX1 2DL To whom it may concern, I am writing to confirm that the thesis submitted by Felix Seifert has been written in correct English throughout its entire text. Yours, Robert Grant Dounton Dr. Robert Grant-Downton Stipendiary Lecturer in Biological Sciences, St. Peter's College, University of Oxford Registered Charity Number 1143166

XIII. Acknowledgements

I would especially like to thank Dr. Stefan Scholten for providing me the chance to conduct my PhD studies on this fascinating project with multi-faceted intersections of developmental biology, bioinformatics and plant breeding. I enjoyed our fruitful discussions resulting in many new ideas and interesting interpretation of results. I furthermore appreciate the willingness to support me at any time if problems or questions arose and providing high performance IT equipment.

I would like to thank Prof. Dr. Arp Schnittger to agree to examine my thesis and the friendly and open affiliation of our lab group.

Furthermore I would like to thank Prof. Dr. Heike Pospisil for providing a position at the TH Wildau and discussions regarding bioinformatical problems.

I want to express my special thanks to my former colleague Dr. Alexander Thiemann for our discussions helping to improve the design of analyses and interpretation of data and proof-reading. Besides our mutually supportive teamwork I got to know you as a good friend and enjoyed our post-conference road trips as well as after work spare ribs.

I would like to thank Dominika Rybka for her excellent technical assistance which was essential for the later outcome of the sequencing analysis. Furthermore I would like to thank all former and current lab members for their support and a nice time together.

I would like to thank Dr. Robert Grant-Downton for discussions, proof-reading of my thesis.

Furthermore I would like to thank for providing material or data required for this thesis: Dr. Junjie Fu for normalized microarray expression data, Dr. José Marquez-Guiterrez for providing the initial P1/HC-Pro plants, and Dr. Tobias Schrag for heterosis field data.

I am deeply grateful to my family and friends supporting and encouraging me throughout the thesis. Finally, I would like to thank you, Friederike, for your unconditional support and enduring both enthusiastic status reports and stressful times and multiple proof-reading.