TranSNIpтомics: Genome-wide transcription profiles provided by NGS-based Massive Analysis of cDNA Ends (MACE) simultaneously identify allele-specific differential expression of root-trait-related drought-response genes in drought-tolerant and susceptible chickpea varieties

Peter Winter, Kamila Bokszczanin, Himabindu Kudapa, Alejandro Rodriguez Meisel, Nicolas Krezdorn, Ruth Jüngling, Rajeev Varshney, Guenter Kahl, Björn Rotter

GenXPro GmbH, Frankfurt am Main

www.genxpro.de
Transcriptome:
- Massive Analysis of cDNA Ends (MACE)
- Bacterial SuperSAGE
- Normalization of cDNA libraries (qualitative information)
- RNA-seq
- Small RNAs / microRNA in tissues, body fluids, exosomes
- Other non-coding RNAs, Degradome
- qPCR service

Genome:
- Whole-genome Sequencing
- Digital karyotyping (ST-DK), RC-seq, CNVs
- Methylation-specific DK (ST-MSDK), Meth-seq
- All Exome sequencing, Target Enrichment

Metagenome:
- COXI, 16s rRNA, others...

Bioinformatics:
- NGS Data Handling, Assembly, Quantification, BLAST
- Expression Data Interpretation, Gene Ontology
### GenXPro: Our Service Portfolio

**Nucleotide-based information**

### Illumina Hiseq2000 vs. PacBio

<table>
<thead>
<tr>
<th></th>
<th>Illumina Hiseq2000</th>
<th>PacBio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequence length</strong></td>
<td>2 x 150 bp</td>
<td>~ 500-15,000 bp (!)</td>
</tr>
<tr>
<td><strong>Throughput/subunit</strong></td>
<td>30-60 Giga Bases</td>
<td>250 Mega Bases</td>
</tr>
</tbody>
</table>

**Full service:** Transcriptomics, Genomics, Genotyping, Epigenomics, **Bioinformatics**

- Patented technique for reduced representation analyses
- Method to eliminate PCR-copies from dataset
- No prior knowledge about NGS required, no hardware, no software, just samples…
How to handle Gigabytes of Data?

Hardware/Computers:
- 164 CPUs, 704 Gigabyte RAM

Assembly:
- different assembly programs available

Annotation:
- Novoalign, BLAST, SOAP, BLAT, Annovar

Enrichment Analysis:
- Gene Ontology, KEGG, BioCarta, GSEA etc.
Drought is the major constraint to chickpea production: A case for TranSNiPtomics

From poster of Manish Roorkiwal, ICRISAT, 2/7/2013
"TranSNiPtomics":

simultaneous analysis of gene expression AND polymorphism = allel-specific gene expression measurement

Advantages:

• Markers located within genes - very likely connected to specific trait
• Markers can be chosen from differentially expressed genes to increase chance of involvement in trait

Requirements:

• Sufficient coverage - distinguish between sequencing error and SNP
• Accurate measurement of transcription levels
Differential gene expression results in large differences of transcript representation in all transcriptomes.

Frequencies of transcript species:
- > 50% of transcripts are present in less than 10 copies
- Less than 0.2% of genes contribute more than 40% of all transcripts
RNA-Seq

- Many reads per transcript
- Reads per transcript vary, depending on transcript length
- Quantification often difficult in non-model organisms
- Very deep sequencing required for short and low-abundant transcripts (e.g. transcription factors, receptors)
Our solution = **Massive Analysis of cDNA Ends (MACE)**

only the cDNA-3‘ends (or 5‘-ends) are sequenced

- Reduced complexity, less variants, but:
- concentration on the **most polymorphic region in a gene**
- highly specific for good annotation!
- easy to quantify!
- high coverage for SNP detection!
- **low costs**!
- **hundreds of genotypes can be analysed**, e.g. mapping populations **at reasonable costs**
Massive Analysis of cDNA Ends (MACE):

How it works
Fragmentation, washing

Massive Analysis of cDNA Ends: MACE

How it works

Streptavidin-Beads

100-300 bp
2nd generation sequencing of 50-100 bp

Massive Analysis of cDNA Ends: MACE

How it works
Massive Analysis of cDNA Ends: MACE

Assembly & Counting

How it works
Massive Analysis of cDNA Ends: MACE

Assembly & Counting

- Green: 1
- Red: 1
- Blue: 1
- Yellow: 4

50-400 bp

Only one fragment per transcript!
Bioinformatics: automated workflow for model and non-model organisms

Tags:
- annotation / mapping
- quantification
  - Gen 1: 1
  - Gen 2: 1
  - unknown: 4

BLASTX (Protein DBs)

Enrichment analysis
RNA-Seq vs. MACE

RNA-Seq

Many reads per transcript, reads per transcript varies!

MACE

one read = one transcript

For similar resolution, RNA-Seq requires about 20-30 times more sequencing*

*Asmann et. al 2009
TranSNiPtomics - why MACE?

High coverage to distinguish between SNP and error

MACE = reduced complexity

Concentration on polymorph 3’ end: SNPs with enough coverage: 2

RNA-Seq = high complexity

Reads distributed all over transcript: SNPs with enough coverage: 0
Sufficient coverage for SNP detection!

MACE, 20 Mio Reads

Wheat, nucleosome/chromatin assembly factor C; 160 TPM
Coverage too low!

RNA seq, 20 Mio reads, same position
IGSTC-Project: IND 09/515
Biotechnological approaches to improve chickpea crop productivity for farming community and industry

Dr Rajeev Varshney
ICRISAT, Patancheru, India

Prof. Dr. Günter Kahl
Molecular BioSciences, Frankfurt University, Germany

Dr. Manash Chatterjee
BenchBio Private Ltd, India

Dr. Peter Winter,
GenXPro GmbH, Germany
Aim:

• Understand the impact and mode of action of a major QTL for drought tolerance present in chickpea variety ICC4958 on drought tolerance in different genetic background

• Identify the genes underlying the QTL

• Produce qRT-PCR markers for transfer of the genes

• Produce transgenics containing the gene(s)
The genomic region around SRR marker TAA170 on chickpea linkage group 4 contains a major QTL for drought tolerance.
Chickpea Drought Research

Genotypes used

ICC4958 (Tolerance Donor)

ICC1882 (drought-susceptible)

\[\text{X} \quad \text{F2} \quad \text{X}\]

129 Recombinant Inbred Lines

Selfing

Tolerant  \hspace{1cm} \text{Susceptible}

JG11 (Indien elite line)

Marker-Assisted Backcrossing

JG11

JG11Plus (High yielding under drought)
### MACE libraries

<table>
<thead>
<tr>
<th>Chickpea libraries</th>
<th>Well Watered - WW:</th>
<th>4 days drought -04:</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC4958-WW</td>
<td>ICC4958-04</td>
<td></td>
</tr>
<tr>
<td>ICC1882-WW</td>
<td>ICC1882-04</td>
<td></td>
</tr>
<tr>
<td>JG11-WW</td>
<td>JG11-04</td>
<td></td>
</tr>
<tr>
<td>JG11plus-WW</td>
<td>JG11plus-04</td>
<td></td>
</tr>
<tr>
<td>RILsS-WW</td>
<td>RILsS-04</td>
<td></td>
</tr>
<tr>
<td>RILsR-WW</td>
<td>RILsR-04</td>
<td></td>
</tr>
</tbody>
</table>

### Reference Transcriptome Databases

- **Chickpea Transcriptome Assembly** (GenXPro, L. Belarmino)
- **C. reticulatum** (PI489777) 27.06.2012; CTDB
- **C. arietinum** (ICC4958) transcriptome; CTDB; Hybrid assembly
- **C. arietinum** (ICC4958) transcriptome; CTDB; Short read assembly
- Unigenes (NCBI) (assembly of ESTs available at NCBI, CTDB
- Refseq plant RNA, downloaded from NCBI database
- *Medicago_sativa_NCBI_Entrez_EST_19082011*
- *Trinity.fasta*
- *all_TIGR_DFCI_PLANT; Dana-Farber Cancer Institute*
Transcript variants (TVs) up- (log2 $\geq 2$, p-value $< 1e^{-3}$) and down-regulated (log2 $\leq -2$, p-value $< 1e^{-3}$) under stress in the different genotypes.

Under drought, ICC4958 regulates many genes up, ICC1882 regulates many genes down.
Gene Expression Profiles of the different Genotypes stressed and well watered

**Priming!**

Already under well-watered conditions gene expression in drought-tolerant ICC4958 clusters with drought-stressed other varieties

Clustered heat-map of gene expression in response to drought stress in roots of susceptible and tolerant chickpea varieties

Blue = well watered
Red = drought stressed
<table>
<thead>
<tr>
<th>High performing Recurrent Parent</th>
<th>NIL</th>
<th>Tolerant parent</th>
<th>Donor Parent</th>
<th>RIL bulks</th>
<th>Susceptible parent</th>
</tr>
</thead>
<tbody>
<tr>
<td>JG11WW JG1104</td>
<td>JG11plusWW JG11plus04</td>
<td>ICC4958WW ICC495804</td>
<td>RILsRWW RILsR04</td>
<td>RILsSWW RILsS04</td>
<td>ICC1882WW ICC188204</td>
</tr>
<tr>
<td>1 regulation of metabolic process</td>
<td>respiratory electron transport chain</td>
<td>respiratory electron transport chain</td>
<td>oxidoreductase activity, acting on NADH or NADPH</td>
<td>NADH dehydrogenase (quinone) activity</td>
<td>oxidation-reduction process</td>
</tr>
<tr>
<td>2 negative regulation of catalytic activity</td>
<td>NADH dehydrogenase activity</td>
<td>NADH dehydrogenase activity</td>
<td>energy derivation by oxidation of organic compounds</td>
<td>NADH dehydrogenase activity</td>
<td>oxidoreductase activity</td>
</tr>
<tr>
<td>3 negative regulation of molecular function</td>
<td>oxidoreductase activity, acting on NADH or NADPH, quinone or similar compound as acceptor</td>
<td>NADH dehydrogenase (ubiquinone) activity</td>
<td>cellular respiration</td>
<td>NADH dehydrogenase (ubiquinone) activity</td>
<td>monoxygenase activity</td>
</tr>
<tr>
<td>4 regulation of catalytic activity</td>
<td>NADH dehydrogenase (quinone) activity</td>
<td>NADH dehydrogenase (quinone) activity</td>
<td>NADH dehydrogenase activity</td>
<td>oxidoreductase activity, acting on NADH or NADPH, quinone or similar compound as acceptor</td>
<td>response to biotic stimulus</td>
</tr>
<tr>
<td>5 regulation of molecular function</td>
<td>NADH dehydrogenase (ubiquinone) activity</td>
<td>oxidative phosphorylation</td>
<td>NADH dehydrogenase (quinone) activity</td>
<td>oxidoreductase activity, acting on NADH or NADPH</td>
<td>iron ion binding</td>
</tr>
<tr>
<td>6 enzyme inhibitor activity</td>
<td>cellular respiration</td>
<td>ATP synthesis coupled electron transport</td>
<td>NADH dehydrogenase (ubiquinone) activity</td>
<td>oxidative phosphorylation</td>
<td>lipid localization</td>
</tr>
<tr>
<td>7 regulation of biological process</td>
<td>oxidoreductase activity, acting on NADH or NADPH</td>
<td>electron transport chain</td>
<td>respiratory electron transport chain</td>
<td>ATP synthesis coupled electron transport</td>
<td>lipid transport</td>
</tr>
<tr>
<td>8 regulation of primary metabolic process</td>
<td>electron transport chain</td>
<td>cellular respiration</td>
<td>oxidoreductase activity, acting on NADH or NADPH, quinone or similar compound as acceptor</td>
<td>respiratory electron transport chain</td>
<td>heme binding</td>
</tr>
<tr>
<td>9 regulation of cellular metabolic process</td>
<td>energy derivation by oxidation of organic compounds</td>
<td>oxidoreductase activity, acting on NADH or NADPH</td>
<td>electron transport chain</td>
<td>electron transport chain</td>
<td>endopeptidase regulator activity</td>
</tr>
<tr>
<td>10 biological regulation</td>
<td>ATP synthesis coupled electron transport</td>
<td>oxidoreductase activity, acting on NADH or NADPH, quinone or similar compound as acceptor</td>
<td>generation of precursor metabolites and energy</td>
<td>protein oligomerization</td>
<td>endopeptidase inhibitor activity</td>
</tr>
<tr>
<td>11 sequence-specific DNA binding transcription factor activity</td>
<td>oxidative phosphorylation</td>
<td>energy derivation by oxidation of organic compounds</td>
<td>oxidative phosphorylation</td>
<td>organelle membrane</td>
<td>peptidase inhibitor activity</td>
</tr>
<tr>
<td>12 nucleic acid binding transcription factor activity</td>
<td>oxidation-reduction process</td>
<td>protein oligomerization</td>
<td>ATP synthesis coupled electron transport</td>
<td>mitochondrial membrane</td>
<td>peptidase regulator activity</td>
</tr>
</tbody>
</table>
Conclusion: The drought-tolerance QTL from ICC4958 is responsible for mitochondrial drought responses of JG11Plus

<table>
<thead>
<tr>
<th>GO Nr.</th>
<th>JG11plus WW</th>
<th>JG11plus-04</th>
<th>ICC4958 WW</th>
<th>ICC4958-04</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>respiratory electron transport chain</td>
<td></td>
<td>respiratory electron transport chain</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>NADH dehydrogenase activity</td>
<td></td>
<td>NADH dehydrogenase (ubiquinone) activity</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>oxidoreductase activity, acting on NADH or NADPH, quinone or similar compound as acceptor</td>
<td></td>
<td>NADH dehydrogenase (quinone) activity</td>
<td>oxidative phosphorylation</td>
</tr>
<tr>
<td>4</td>
<td>NADH dehydrogenase (quinone) activity</td>
<td></td>
<td></td>
<td>ATP synthesis coupled electron transport</td>
</tr>
<tr>
<td>5</td>
<td>NADH dehydrogenase (ubiquinone) activity</td>
<td></td>
<td></td>
<td>electron transport chain</td>
</tr>
<tr>
<td>6</td>
<td>cellular respiration</td>
<td></td>
<td></td>
<td>cellular respiration</td>
</tr>
<tr>
<td>7</td>
<td>oxidoreductase activity, acting on NADH or NADPH</td>
<td></td>
<td></td>
<td>oxidoreductase activity, acting on NADH or NADPH</td>
</tr>
<tr>
<td>8</td>
<td>electron transport chain</td>
<td></td>
<td></td>
<td>quinone or similar compound as acceptor</td>
</tr>
<tr>
<td>9</td>
<td>energy derivation by oxidation of organic compounds</td>
<td></td>
<td></td>
<td>oxidoreductase activity, acting on NADH or NADPH</td>
</tr>
<tr>
<td>10</td>
<td>ATP synthesis coupled electron transport</td>
<td></td>
<td></td>
<td>quinone or similar compound as acceptor</td>
</tr>
</tbody>
</table>
RNA from Roots of:

ICC4958 (tolerant) x ICC1885 (susceptible)

Recombinant Inbred Lines (RILs)

Tolerant bulk: 9 best

Susceptible bulk: 9 worst
### Number of sequenced and annotated MACE tags from roots of ICC4958, ICC1882 and the bulks

<table>
<thead>
<tr>
<th>Library</th>
<th>Total no of annotated tags</th>
<th>No of tags not annotated</th>
<th>Annotated in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC4958_WW</td>
<td>9‘867‘301</td>
<td>3‘734‘104</td>
<td>62,16</td>
</tr>
<tr>
<td>ICC4958_04</td>
<td>12‘373‘412</td>
<td>3‘226‘189</td>
<td>73,93</td>
</tr>
<tr>
<td>ICC1882_WW</td>
<td>9‘422‘171</td>
<td>2‘269‘180</td>
<td>75,92</td>
</tr>
<tr>
<td>ICC1882_04</td>
<td>7‘364‘703</td>
<td>2‘532‘140</td>
<td>65,62</td>
</tr>
<tr>
<td>Tolerant Bulk_WW</td>
<td>11‘066‘463</td>
<td>4‘620‘331</td>
<td>58,25</td>
</tr>
<tr>
<td>Tolerant Bulk_04</td>
<td>11‘902‘727</td>
<td>5‘160‘686</td>
<td>56,64</td>
</tr>
<tr>
<td>SusceptibleBulk_WW</td>
<td>38‘714‘730</td>
<td>10‘496‘416</td>
<td>72,89</td>
</tr>
<tr>
<td>SusceptibleBulk_04</td>
<td>27‘310‘911</td>
<td>4‘494‘668</td>
<td>83,54</td>
</tr>
</tbody>
</table>

**Blue = well watered**

**Red = 4 days drought stressed**
Results of Bulked Segregant Analysis
Differentially regulated genes

Susceptible RILs regulate twice as many genes down as Resistant RILs

Well Watered
- Tolerant
  - Susceptible
  - 177 downregulated transcripts
  - 277 upregulated transcripts

Drought
- Tolerant
  - Susceptible
  - 304 downregulated transcripts
  - 447 upregulated transcripts

- Resistant
  - Susceptible
  - 707 downregulated transcripts
  - 479 upregulated transcripts

P value ~0
- down regulated transcripts
- up regulated transcripts

MACE libraries: 4x Parents, 4 x Bulks
Results of Bulked Segregant Analysis
Differentially regulated genes

Susceptible RILs reaction to water stress

Stressed

Well watered
Tolerant RILs reaction to water stress: much less...

Results of Bulked Segregant Analysis
Differentially regulated genes
Tolerant RILs compared to Susceptible RILS under water stress

Results of Bulked Segregant Analysis
Differentially regulated genes
GO Terms most enriched under water deficit in tolerant varieties are strongly related to mitochondrial function.

The crucial role of plant mitochondria in orchestrating drought tolerance
Owen K. Atkin and David Macherel; 2009
Transcripts and alleles that are unique in the tolerant RILs are potentially powerful markers for drought resistance.

58 Transcripts were exclusively found in the tolerant Bulks (>50 copies) under stress and well watered conditions.

Among them:
- heat shock proteins
- LEA proteins
- many unknown or unknown in context of drought
Allele distribution: Whose alleles went where?

ICC4958 (tolerant) × ICC1882 (susceptible)

Recombinant Inbred Lines

Tolerant bulk

Susceptible bulk
Automated bioinformatics workflow for model and non-model organisms

1) annotation to reference or to de novo assembly of MACE tags of all libraries

2) SNP detection: Position of SNP and sequence 100 bp before and after SNP are provided for primer design

Excerpt of output table:

<table>
<thead>
<tr>
<th>Transcript</th>
<th>position</th>
<th>ref_base</th>
<th>var_base</th>
<th>Library I ref base</th>
<th>Library I var base</th>
<th>100 bp before SNP</th>
<th>SNP</th>
<th>100 bp behind SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>comp12617_c</td>
<td>54</td>
<td>A</td>
<td>G</td>
<td>33</td>
<td>0</td>
<td>AGCCAGGCTCTCTGC</td>
<td>A/G</td>
<td>ATTCGGAGTGAGGTCGACCA'</td>
</tr>
<tr>
<td>comp12617_c</td>
<td>119</td>
<td>C</td>
<td>T</td>
<td>125</td>
<td>0</td>
<td>TGGTATGCAGGCTTTT</td>
<td>C/T</td>
<td>GAGTTGTCGCTTTTATTTTA</td>
</tr>
<tr>
<td>comp14620_c</td>
<td>169</td>
<td>T</td>
<td>C</td>
<td>194</td>
<td>2</td>
<td>GACAAGGTTAGGAACCT</td>
<td>T/C</td>
<td>AAGTGTTGCTAAATAATCTTA</td>
</tr>
<tr>
<td>comp200644c</td>
<td>60</td>
<td>C</td>
<td>T</td>
<td>0</td>
<td>11</td>
<td>GTATACTTTTATGTA</td>
<td>C/T</td>
<td>TTCTTTCAATTTTTATCCAT</td>
</tr>
<tr>
<td>comp20869_c</td>
<td>132</td>
<td>T</td>
<td>G</td>
<td>7</td>
<td>159</td>
<td>CGATAAGGCTCTACCG</td>
<td>T/G</td>
<td>CGTGTATATATACTGTATAAA'</td>
</tr>
<tr>
<td>comp20869_c</td>
<td>147</td>
<td>T</td>
<td>A</td>
<td>5</td>
<td>222</td>
<td>CATCACCTCCGCAGC</td>
<td>T/A</td>
<td>TATATAATATTCATTGAGT</td>
</tr>
<tr>
<td>comp21319_c</td>
<td>297</td>
<td>C</td>
<td>T</td>
<td>243</td>
<td>0</td>
<td>ATGCTGCGTGCGAGC</td>
<td>C/A</td>
<td>AAGCATTTTTTTCTCTCTTAT</td>
</tr>
</tbody>
</table>
comp21149_c0_seq1 = not annotated
comp21149_c0_seq1 = not annotated
Summary:
Alleles detected in parental lines and tolerant and susceptible bulks

Parents (ICC4958 and ICC 1885):

Alleles, at least 10 x exclusively found in either one parent: 3896
Contrasting, exclusive alleles, each allele at least 10 x in both parents 128

Susceptible and tolerant Bulks:

Alleles, at least 10 x exclusively found in either one bulk: 1234
Contrasting, exclusive alleles expressed at least 10 x in both bulks: 12
In the resistant bulk transcription profiles ICC4958 alleles are strongly over-represented.
TranSNiPtomics: SNPs in the QTL on LG4

Whose genes went where?
TranSNiPtomics: SNPs in genes in the LG4 QTL region

Genomic Position: Ca4_13,726,396_13,728,300
Gene: Ninja-family protein mc410

Polymorphic base Position in gene
G 322  T 322  A 1220  C 1220  T 1390  C 1390

Whose genes went where?
MACE detected 30 SNPs in 11 genes in an important 166,084kbp long region of the chickpea genome

TranSNiPtomics: SNPs in genes in the LG4 QTL region

<table>
<thead>
<tr>
<th>Gene</th>
<th>Genomic Position</th>
<th># of SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cysteine synthase-like</td>
<td>Ca4_13,678,141_13,678,617</td>
<td>4</td>
</tr>
<tr>
<td>FRIGIDA-like protein</td>
<td>Ca4_13,684,591_13,686,920</td>
<td>6</td>
</tr>
<tr>
<td>FRIGIDA-like protein</td>
<td>Ca4_13,688,248_13,689,022</td>
<td>3</td>
</tr>
<tr>
<td>Suppressor of gene silencing 3 homolog</td>
<td>Ca4_13,693,728_13,695,332</td>
<td>3</td>
</tr>
<tr>
<td>Uncharacterized LOC101489729</td>
<td>Ca4_13,699,049_13,700,445</td>
<td>2</td>
</tr>
<tr>
<td>Ninja-family protein mc410</td>
<td>Ca4_13,726,396_13,728,300</td>
<td>3</td>
</tr>
<tr>
<td>Uncharacterized LOC101494058</td>
<td>Ca4_13,768,316_13,768,909</td>
<td>1</td>
</tr>
<tr>
<td>Primary amine oxidase</td>
<td>Ca4_13,788,285_13,789,046</td>
<td>2</td>
</tr>
<tr>
<td>Uncharacterized LOC101495327</td>
<td>Ca4_13,797,642_13,799,515</td>
<td>2</td>
</tr>
<tr>
<td>TIME FOR COFFEE-like</td>
<td>Ca4_13,835,369_13,839,221</td>
<td>2</td>
</tr>
<tr>
<td>TIME FOR COFFEE-like</td>
<td>Ca4_13,843,038_13,844,225</td>
<td>2</td>
</tr>
</tbody>
</table>
Ongoing Experiment:

MACE analysis of the stress responses of 75 RILs from the cross ICC4958 x 1882

Sequencing performed so far

<table>
<thead>
<tr>
<th></th>
<th>ICC4958 Parent</th>
<th>Tolerant RILs</th>
<th>ICC1882 Parent</th>
<th>Susceptible RILs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># Individuals</td>
<td>2</td>
<td>41</td>
<td>2</td>
<td>33</td>
<td>78</td>
</tr>
<tr>
<td># Reads (Mio)</td>
<td>17.8</td>
<td>362.8</td>
<td>7.0</td>
<td>274.4</td>
<td>662.1</td>
</tr>
</tbody>
</table>
• GO-enrichment analysis of drought-responsive transcripts reveals strong regulation of respiratory transport chain-transcripts.

• Besides typical drought-responders like dehydrin-1, LEA-proteins and heat shock proteins, many currently un-described reactive transcripts were identified.

• New, highly reliable SNPs /alleles in the drought tolerant RIL-bulks

• MACE = cost-efficient, simultaneous gene expression and genotyping !

• Just started: Analysis of 75 RILs with MACE

• MACE-Kit will be released this year.
Thank you for your patience and:

Himabindu Kudapa
Rajeev Kumar Varshney

Nicolas Krezdorn
Björn Rotter
Kamila Bokszczanin
Anja Frank
Peter Winter

Jutta Kreutz
Nicolas Gonzales
Günter Kahl

The Indogerman Program (BMBF)
Thank you!
Breeding lines used for TranSNiPtomics

A) At ICRISAT, the major QTL for drought tolerance was transferred from ICC4958 (drought tolerant, DP) via marker-assisted back crossing (MABC) to JG11 (elite line, moderately drought tolerant, RP) to give the NIL JG11Plus (highly drought tolerant, high yielding under drought).

B) ICC4958 (drought tolerant, DP) was crossed to ICC1882 (drought-susceptible). The F2-offspring was self-pollinated to give tolerant and susceptible RILs, RILsS and RILsR, respectively.
Aims:

- Identification markers and/or genes for drought resistance
- Better understanding of drought resistance in roots

Approach

- Massive analysis of cDNA Ends (MACE), a reduced complexity transcriptome sequencing method for simultaneous gene expression analysis and genotyping
Transcriptomics of Thermotolerance

Transcript annotation

MACE library prep

- Clustering 94 bp sequences with 100% of identity
- Mean quality score for each bp
- MACE tags frequencies with quality scores
- polyA trimming

Reference Transcriptome Databases

- Redundancy

Reference Protein Repository Databases

- NR
- UniprotKB/Swissprot
- Mt3.5

formatdb
Transcriptomics of Thermotolerance

Transcript annotation

MACE library prep

- clustering 94 bp sequences with 100% of identity
- Mean quality score for each bp
- MACE tags frequencies with quality scores
- polyA trimming

Reference Transcriptome Databases

Reference Protein Repository Databases

BLASTX

- take ids
- combine ids

combine blastx tables in the order:
- nr, swissprot, Mt3.5
MACE library prep

Library

clustering 94 bp sequences with 100% of identity

Mean quality score for each bp

MACE tags frequencies with quality scores

polyA trimming

BLASTX

Reference Transcriptome Databases

Reference Protein Repository Databases

NR

Swissprot

Redundant

Mt3.5

formatdb

take ids

combine ids

combine blastx tables in the order: nr, swissprot, Mt3.5

polyA trimming

MACE tags frequencies with quality scores

Mean quality score for each bp

SOAP
Trancriptomics of Thermotolerance
Functional Analysis

GO analysis
Bioinformatics: automated workflow for model and non-model organisms

Tags:
- annotation / mapping
- BLASTX (Protein DBs)

Quantification:
- Gen 1: 1
- Gen 2: 4

Enrichment analysis

WEB tool: "MACE2GO"
“TranSNIPTomics”:

simultaneous analysis of gene expression AND polymorphisms;
advantages:

- Markers located within genes - very likely connected to specific trait
- Markers can be chosen from differentially expressed genes to increase chance of involvement in trait

Requirements:

- Sufficient coverage - distinguish between sequencing error and SNP
- Accurate measurement of transcription levels
Experimental setup

MACE libraries: 4x Parents, 4 x Bulks