Plant Physiology Preview. Published on April 13, 2010, as DOI:10.1104/pp.110.156448

Running head: Mutation detection in non-reference genomes

Keywords: *Arabidopsis thaliana*, spontaneous mutation, SQE1, whole genome sequencing, next generation sequencing

Corresponding author:
Detlef Weigel
Max-Planck-Institute for Developmental Biology
Spemannstraße 37-39
72076 Tübingen
Germany
Phone: +49-(0)7071-601 1411
Fax: +49-(0)7071-601 1412
Email: weigel@weigelworld.org

Scientific Correspondence

Identification of a spontaneous frame shift mutation in a non-reference *Arabidopsis thaliana* accession using whole genome sequencing[1]

Roosa A. E. Laitinen, Korbinian Schneeberger, Noémie S. Jelly[2], Stephan Ossowski, and Detlef Weigel*

Department of Molecular Biology, Max Planck Institute for Developmental Biology, D-72076 Tübingen, Germany.

[1] This work was supported by a Human Frontier Science Program Long-Term Fellowship (R.A.E.L.), a Gottfried Wilhelm Leibniz Award of the Deutsche Forschungsgemeinschaft and the Max Planck Society.

[2] Present address: Laboratoire Vigne Biotechnologies et Environnement, 68008 Colmar Cedex, France

* Corresponding author; e-mail weigel@weigelworld.org

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Detlef Weigel (weigel@weigelworld.org).

Short-read sequencing technologies support the facile de novo identification of spontaneous and chemically induced mutations in the *Arabidopsis thaliana* reference genome (Schneeberger et al., 2009; Ossowski et al., 2010). Here, we show that short read sequencing is also suitable for the analysis of new mutations in a non-reference inbred accession that differs from the reference genome in about 0.5% of all positions.

We crossed two normal appearing, green individuals of *Arabidopsis thaliana* accessions, Krotzenburg (Kro-0, CS1301) and Anholt (Anh-1, CS22313) to each other. The F_1 plants were all normal, but the F_2 population segregated purplish, small and non-flowering plants (Fig. 1A). Plants could be prompted to flower in high humidity, but the resulting seeds were not viable (Fig. 1B). Leaves were about 10 times smaller than in wild type, but leaf cell number was reduced only about three fold, indicating that both decreased cell expansion and division contributed to the dwarf phenotype. Consistent with the purplish phenotype, several genes involved in biosynthesis of the purple pigment anthocyanin were upregulated in the dwarf plants. Using a combination of per-gene variance (rank product p-value of 0.01) (Breitling et al., 2004) and common variance (two-fold change) as criterion, *PRODUCTION OF ANTHOCYANIN PIGMENT 2* (*PAP2*; At1g66390), *CHALCONE ISOMERASE* (*CHI*; At3g55120), *FLAVONOL SYNTHASE 1* (FLS1; At5g08640), and *DIHYDROFLAVONOL 4-REDUCTASE* (*DFR*; At5g42800) were all differentially expressed. None of these phenotypes, including a disorganized root (Fig. 1C, D), could be suppressed by treating the plants with cytokinin, gibberellic acid, jasmonic acid or auxin.

Using conventional mapping with almost 1,900 F_2 plants of the Kro-0 x Anh-1 cross, we identified a 530 kb interval, between 21.36 and 21.88 Mb on chromosome 1, that was linked to the dwarf phenotype (Fig. 2). The mapping interval contained 116.5 kb of repetitive DNA, which is often polymorphic and may suppress recombination (Fu et al., 2002), possibly explaining the failure to further reduce the final mapping interval.

Based on the sequences of the flanking markers, we concluded that plants showed the dwarf phenotype, if they had inherited both alleles from the Kro-0 grandparent used in the cross to Anh-1. Since the original Kro-0 line did not exhibit the dwarf phenotype, and other Kro-0 x Anh-

1 crosses did not produce abnormal F_2 progeny, we concluded that a spontaneous mutation had occurred in the germline of the particular Kro-0 individual used for the original cross to Anh-1. The F_1 plant would have been heterozygous for this mutation. We therefore decided to directly compare the mutant genome in this interval with that of the Kro-0 parental genome. Because the size of the final mapping interval made analysis by PCR based sequencing impractical, we sequenced the entire Kro-0 parental genome at 25-fold coverage, with 36 to 42 bp paired-end reads generated on Illumina's Genome Analyzer. In parallel, we produced 25-fold coverage of the haploid genome from F_3 dwarf plants. We pooled genomic DNA from 100 plants to obtain sufficient material for sequencing. SNPs and indels were called for both the parent and mutant pool, by independently comparing them to the Col-0 reference genome using SHORE and GenomeMapper (Ossowski et al., 2008; Schneeberger et al., 2009). For background cleaning we made use of all variants detected in the Kro-0 parent. To predict mutations private to the dwarf sample, only those with a SHORE quality value of at least 25 were considered.

Within the 530 kb mapping interval, we identified 5,691 single nucleotide differences in the dwarf pool relative to the Col-0 reference sequence. Of these, 4,023 were predicted with high confidence. This level of polymorphism is similar to that found in other accessions in this region, with 4,036 and 3,511 found in the genomes of Bur-0 and Tsu-1, respectively (Ossowski et al., 2008). Of the 4,023 high-quality polymorphisms, 531 were predicted to change the coding potential of 63 genes. All but one were shared with the normal Kro-0 parent. The one remaining mutation in the dwarf pool, a 1-bp deletion, resided in the seventh exon of the gene At1g58440, located in the middle of the mapping interval at 21.718 Mb. The deletion disrupted the At1g58440 open reading frame (Fig. 2). Dideoxy sequencing confirmed that the mutation was specific to F_3 individuals with the dwarf phenotype. A Col-0 line with a T-DNA insertion in At1g58440 (N522763) showed the same purplish, dwarf and abnormal root phenotype as these plants. At1g58440 encodes SQUALENE EPOXIDASE 1 (SQE1), which catalyzes a key step in sterol metabolism, and the morphological phenotypes of *sqe1* mutants are very similar to the ones seen in our dwarfs, including partial rescue by growing plants in 90% humidity (Rasbery et al., 2007; Posé et al., 2009) (D. Posé, personal communication).

Our study provides a proof of concept for identifying mutations in a background other than a high-quality reference genome using direct whole genome sequencing. We have recently shown that de novo mutation identification with short-read sequencing in a reference background provides not only very high specificity (i.e., very few false positives), but also very high sensitivity (i.e., very few false negatives) (Ossowski et al., 2010). This is in contrast to similar efforts with human genomes (Lupski et al., 2010), reflecting both the more complex nature of human genomes, but also the absence of a near-isogenic reference. Moreover, different from other studies aimed at identifying causal mutations for human diseases (Lupski et al., 2010), we took an unbiased approach in the current work, and did not use any prior information on candidate genes associated with the phenotype in question. In summary, our work indicates that short-read sequencing is a useful and sensitive tool that can be applied to mutation identification, as long as a high-quality reference sequence from close relatives is available. This prospect should be good news for anybody interested in performing mutant screens in non-model organisms.

ACKNOWLEDGMENTS

The authors thank Jun Cao and Christa Lanz for help with Illumina sequencing; Kirsten Bomblies for the F_1 seeds; Kirsten Bomblies, David Posé, Ignacio Rubio-Somoza and Marco Todesco for sharing ideas and discussions; and Waldemar Hauf for technical help.

LITERATURE CITED

Breitling R, Armengaud P, Amtmann A, Herzyk P (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett **573:** 83-92

Fu H, Zheng Z, Dooner HK (2002) Recombination rates between adjacent genic and retrotransposon regions in maize vary by 2 orders of magnitude. ProcNatl Acad Sci USA **99:** 1082-1087

Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, Bainbridge M, Dinh H, Jing C, Wheeler DA, McGuire AL, Zhang F, Stankiewicz P, Halperin JJ, Yang C, Gehman C, Guo D, Irikat RK, Tom W, Fantin NJ, Muzny DM, Gibbs RA (2010) Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. N Engl J Med**:** published online Mar 10, 2010

Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D (2008) Sequencing of natural strains of *Arabidopsis thaliana* with short reads. Genome Res **18:** 2024-2033

Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D (2008) Sequencing of natural strains of *Arabidopsis thaliana* with short reads. Genome Res **18:** 2024-2033

Ossowski S, Schneeberger K, Lucas-Lledo JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science **327:** 92-94

Posé D, Castanedo I, Borsani O, Nieto B, Rosado A, Taconnat L, Ferrer A, Dolan L, Valpuesta V, Botella MA (2009) Identification of the Arabidopsis *dry2*/*sqe1-5* mutant reveals a central role for sterols in drought tolerance and regulation of reactive oxygen species. Plant J **59:** 63-76

Rasbery JM, Shan H, LeClair RJ, Norman M, Matsuda SP, Bartel B (2007) Arabidopsis thaliana squalene epoxidase 1 is essential for root and seed development. J Biol Chem **282:** 17002-17013

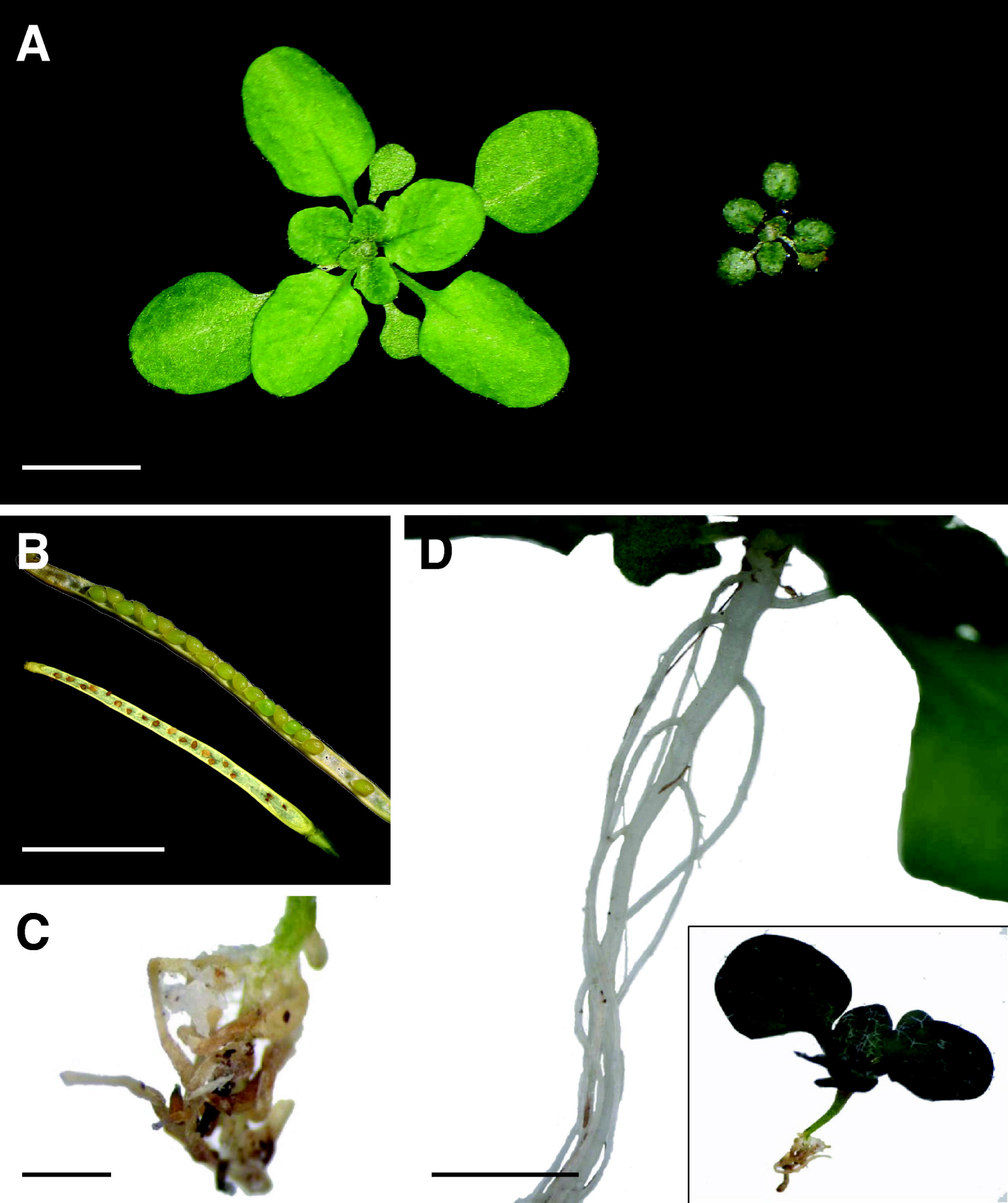
Schneeberger K, Hagmann J, Ossowski S, Warthmann N, Gesing S, Kohlbacher O, Weigel D (2009) Simultaneous alignment of short reads against multiple genomes. Genome Biol **10:** R98

Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jørgensen JE, Weigel D, Andersen SU (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat Methods **6:** 550-551

FIGURE LEGENDS

Figure 1. (A) A quarter of F_2 plants from a single Anh-1 x Kro-0 cross were purplish dwarfs (right), compared to larger, green siblings. (B) Rescue of the flowering defect by spraying plants with water every other day allowed fertilization, but did not support normal seed development (bottom, compared to normal silique above). (C) Close-up of abnormal root of soil-grown plants, with ectopic outgrowths. (D) Comparison of abnormally small root system (inset) of dwarf plants compared to normal siblings, shown at same scale. Scale bars represent 1 cm in B, D, and 0.1cm in C.

Figure 2. Mapping interval (purple) on chromosome 1, and polymorphisms in the vicinity of the causal mutation (red). Green and blue lines indicate single nucleotide changes and deletions, respectively, shared with the parental Kro-0 strain. Bottom shows alignments of Illumina DNA sequence reads against the reference genome sequence, positions 21,714,424 to 21,714,504 (TAIR9). The amino acid sequence encoded by the reverse strand is given below.


A
B
C
D


21.0
22.0 Mb
AT1G58420
AT1G58430
AT1G58440
AT1G58450
AT1G58460
21.71
21.72
Mb
mutant F3 pool
Kro-0 wild type
CCGAAAGCAAAGATACTGGTCCACTTGTGCACATACCCCC-AG
CCGAAAGCAAAGATACTGGTCCACTTGTGCACATACCCCC-AG
CCGAAAGCAAAGATACTGGTCCACTTGTGCACATCCCCCC-AG
CGAAAGCAAAGATACTGGTCCACTTGTGCACATACCCCC-A
GAAAGCAAAGATACTGGTCCACTTGCGCACATACCCCC-A
AAAGCAAAGATACTGGTCCACTTGTGCACATACCCCC-AGGCC
AAGATACTGGTCCACTTGTGCACATACCCCC-AGGCCCAGATA
GATACTGGTCCACTTGTGCACATACCCCC-AGGCCCAG
TGGTCCACTTGTGCACATACCCCC-AGGCCCAGATAATCGAA
TCCACTTGTGCACATACCCCC-AGGCCCAGATAATCG
CACTTGTGCACATACCCCC-AGGCCCAGATAATCGAAGCA
CACTTGTGCACATACCCCC-AGGCCCAGATAATCGAAGCAAGC
ACTTGTGCACATACCCCC-AGGCCCAGATAATCGAAGCAA
CTTGTGCCCATACCCCC-AGGCCCAGATAATCGAAGCAAGC
TTGTGCACATACCCCC-AGGCCCAGATAATCGAAGCAAGCTTC
CACATACCCCC-AGGCCCAGATAATCGAAGCAAGCTTCCC
CACATACCCCC-AGGCCCAGATAATCGAAGCAAGCTTCCCTC
CACATACCCCC-AGGCCCAGATAATCGAAGCAAGCTTCCCTCA
ACATACCCCC-AGGCCCAGATAATCGAAGCAAGCTTCCCTCAT
CCCC-AGGCCCAGATAATCGAAGCAAGCTTCCCTCATCTCGTT
CC-AGGCCCAGATAATCGAAGCAAGCTTCCCTCATCTCGTTTC
C-AGGCCCAGATAATCGAAGCAAGCTTCCCTCATCTCGT
AAAGCAAAGATACTGGTCCACTTGTGCACATACCCCCGAGGC
GCAAAGATACTGGTCCACTTGTGCACATACCCCCGAGGCCCA
CAAAGATACTGGTCCACTTGTGCACATACCCCCGAGGCCCAG
AAGATACTGGTCCACTTGTGCACATACCCCCGAGGCCCAGA
GTGCTGGTCCACTTGTGCACATACCCCCGAGGCCCA
TACTGGTCCACTTGTGCACATACCCCCGAGGCCCAGATAATC
ACTGGTCCACTTGTGCACATACCCCCGAGGCCCAGATAATC
TGGACCACTTGTGCACATACCCCCGAGGCCCAGATAATCGA
TGGTCCACTTGTGCACATACCGCCGAGGCCCAGATAAT
GGTCCAATTGTGCACATACCCCCAGGCCCCAGATAATCGA
GTCCACTTGTGCACATACCCCCGAGGCCCAGATAATCGAAGC
TCCACTTGTGCACATACCCCCGAGGCCCAGATAATCG
ACTTGTGCACATACCCCCGAGGCCCAGATAATCGAA
ACTTGTGCACATACCCCCGAGGCCCAGATAATCGAAGCAAGC
CTTGTGCACATACCCCCGAGGACCCGATAATCGAAGCAAG
TGTGCACATACCACCGAGGCCCAGATAATCGAAGCA
GCACATACCCCCGAGGCCCAGATAATCGAAGCAAGC
GCACATACCCCCGAGGCCCAGATAATCGAAGCAAGC
CACATACCCCCGAGGCCCAGATAATCGAAGCAAGCTTCCCTC
CCCCCGAGGCCCAGATAATCGAAGCAAGCTTCCCTCATCTC
CCCCCGAGGCCCAGATAATCGAAGCAAGCTTCCCTCATCTCG
CCCCCGAGGCCCAGATAATCGAAGCAAGCTTCCCTCATCTCG
CCCAAGGCCCAGATAATCGAAGCAAGCACTCCCTCATCTCGTT
CCGAGGCCCAGATAATCGAAGCAAGCTTCCCTCATCTCGTT
GAGGCCCAGATAATCGAAGCAAGCTTCCCTCATCTCGTTTC
CCGAAAGCAAAGATACTGGTCCACTTGTGCACATACCCCCGAGGCCCAGATAATCGAAGCAAGCTTCCCTCATCTCGTTTC
< S L L S V P G S T C M G G L G L Y D F C A E R M E N R