# Technoreview

# SuperSAGE

**Hideo Matsumura,[1] Akiko Ito,[1] Hiromasa Saitoh,[1,2] Peter Winter,[3,4] Günter Kahl,[3,4] Monika Reuter,[5] Detlev H. Krüger[5] and Ryohei Terauchi[1]***
[1]*Iwate Biotechnology Research Center, 22-174-4, Narita, Kitakami, Iwate 024-0003, Japan.*
[2]*Max-Planck-Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, D-50829, Cologne, Germany.*
[3]*Plant Molecular Biology, Biocenter, University of Frankfurt, Marie-Curie-Strasse 9, D-60439, Frankfurt am Main, Germany.*
[4]*GenXPro, Frankfurt Innovation Centre Biotechnology, Altenhöfer Allee 3, D-60438, Frankfurt am Main, Germany.*
[5]*Institut für Virologie, Universitätsklinikum Charité, Humboldt University, D-10098 Berlin, Germany.*

## Summary

**The application of transcriptomics to study host–pathogen interactions has already brought important insights into the mechanisms of pathogenesis, and is expanding further keeping pace with the accumulation of genomic sequences of host organisms (human and economically important organisms such as food crops) and their pathogens (viruses, bacteria, fungi and protozoa). In this review, we introduce Super-SAGE, a substantially improved variant of serial analysis of gene expression (SAGE), as a potent tool for the transcriptomics of host–pathogen interactions. Notably, the generation of 26 bp tags in the Super-SAGE procedure allows to decipher the 'interaction transcriptome', i.e. the simultaneous monitoring of quantitative gene expression, of both a host and one of its eukaryotic pathogens. The potential of Super-SAGE tags for a rapid functional analysis of target genes is also discussed.**

## The analysis of gene expression during host–pathogen interactions

Infectious diseases cost lives of over 14 million people

each year (WHO, 2003). Similarly, diseases reduce crop production by more than 10% worldwide (ISPP, 1998). Therefore, it is of primary importance to effectively control infectious diseases both from medical and agricultural points of view. The study of host–pathogen interactions is instrumental for such purposes. Host eukaryotes are constantly exposed to attacks by microbes seeking to colonize and propagate in host cells. To counteract them, host cells utilize a whole battery of defence systems to combat microbes. However, in turn, successful microbes evolved sophisticated systems to evade host defence. As such, interactions between hosts and pathogens are perceived as evolutionary arms races between genes of the respective organisms (Bergelson *et al.*, 2001; Kahn *et al.*, 2002; Woolhouse *et al.*, 2002). Any interaction between a host and its pathogen involves alterations in cell signalling cascades in both partners, that may be mediated by transcriptional or post-translational changes. Here, the major challenge for researchers is how to select target genes to be studied in detail from among thousands of genes encoded in the genome. Transcriptomics is one of the methodologies to serve this purpose. Analytical techniques for transcriptomics include differential display (DD; Liang and Pardee, 1992), cDNA-AFLP (Bachem *et al.*, 1996), random EST sequencing (Kamoun *et al.*, 1999), microarray (Schena *et al.*, 1995), serial analysis of gene expression (SAGE; Velculescu *et al.*, 1995) and massively parallel signature sequencing (MPSS; Brenner *et al.*, 2000). Among them, microarray is recently used more frequently than other platforms. Several excellent reviews are available for the use of micoarray for studying host–pathogen interactions (Cummings and Relman, 2000; Diehn and Relman, 2001; Kato-Maeda *et al.*, 2001; Wan *et al.*, 2002; Bryant *et al.*, 2004). In this context it is quite remarkable that most of the gene expression studies addressing host–pathogen interactions in reality examined either host or pathogen separately. However, the simultaneous monitoring of gene expression of both host and pathogen ('interaction transcriptome'), preferably during the infection process and *in situ*, especially in the field of plant–microbe interactions, is at stakes and has already been advocated by Birch and Kamoun (2000). Also in our mind, this approach is necessary to elucidate the host–

pathogen interplay in molecular detail, although presently available techniques cannot discriminate between the transcriptomes of both organisms. In this article, we present a novel method called SuperSAGE, which has proven potential for an analysis of the interaction transcriptome.

## SAGE

Serial analysis of gene expression as developed by Velculescu *et al.* (1995) is a high-throughput method to determine the absolute abundance of every transcript in a population of cells (Fig. 1). Messenger RNAs isolated from cells are converted to double-stranded DNA. After digestion with a 4 bp cutter 'anchoring enzyme' NlaIII, the poly-A proximal ends are collected and ligated to a linker fragment. This linker fragment harbours a 5′-GGGAC-3′ sequence, which is the recognition site of the Type IIS restriction endonuclease BsmFI that cleaves the cDNA 15 bp away in the 3′ direction from the recognition site. Treatment of the linker-ligated cDNA with BsmFI therefore releases a 15 bp fragment called 'tag' from a defined position of each cDNA. After removal of the linker fragment, tags are concatenated, cloned into a plasmid vector, and sequenced. Usually, 10 000–100 000 tags are analysed in total for a given sample. The numbers of each tag in the total sample (tag count) faithfully represent the abundance of the transcript corresponding to the tag. The next step is to identify the gene corresponding to the tag (tag annotation). The 15 bp tag sequence is used as query to search expressed sequence tags (ESTs) or cDNA databases of the organism of interest by BLAST (Altschul *et al.*, 1990). Results of tag counts and tag annotation are finally combined into a gene expression profile. By comparing gene expression profiles of two samples that are differently treated, we can tell which gene is up- or downregulated in response to the particular treatment. SAGE has been widely applied to profile gene expression in yeast (Velculescu *et al.*, 1997), humans (Polyak *et al.*, 1997; Zhang *et al.*, 1997; Polyak and Riggins, 2001; Patino *et al.*, 2002), protozoa (Cummings and Wirth, 2001; Patankar *et al.*, 2001), plants (Matsumura *et al.*, 1999, 2003a) and pathogenic fungi (Steen *et al.*, 2002; Thomas *et al.*, 2002; Irie *et al.*, 2003). In contrast to the analogue data set generated by microarray analysis capitalizing on the relative strength of the hybridization signal of each spot, SAGE data, i.e. transcript counting, is digital and suitable for a comparison of different data sets (see Sagemap at NCBI: http://www.ncbi. nlm.nih.gov/SAGE/) and for bioinformatics in general. Furthermore, in any microarray experiment, expression changes can be studied only for the genes spotted on the array, whereas SAGE theoretically can discover and address all the transcripts, so that SAGE can be regarded as an 'open architecture' technique. Another advantage of SAGE over microarray is that SAGE does not require special and costly equipments (such as DNA spotting machines and microarray reader). SAGE can be performed in any laboratory equipped with basic molecular biology facilities. This low requirement of equipment in SAGE procedure contrasts with MPSS (Brenner *et al.*, 2000), another powerful, but facility-demanding high-throughput expression profiling technique. On the other hand, the disadvantage of SAGE is its high requirement of human work and time. For the analysis of two samples and starting from isolated RNA, 10–14 days are required
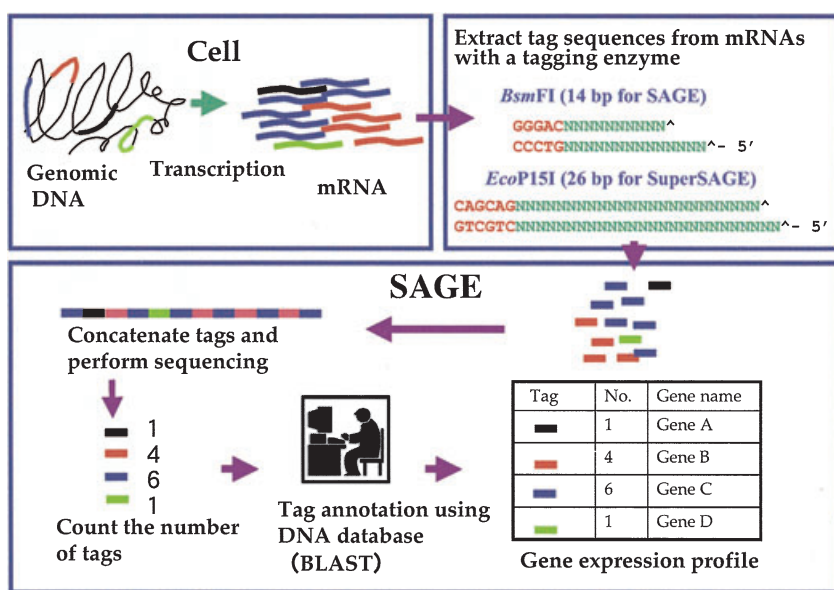


**Fig. 1.** A simplified scheme of SAGE and SuperSAGE. From a defined position of every mRNA of the cells, a short fragment called SAGE 'tag' is isolated by a 'tagging enzyme'. In the SuperSAGE protocol, a Type III restriction endonuclease EcoP15I is used instead of BsmF1 as the tagging enzyme, resulting in the isolation of 26 bp tags. Tags are concatenated and sequenced. Frequency of each tag in the sample (tag count) represents abundance of the transcript corresponding to the tag. Tag sequence is used as BLAST query to identify the gene of origin (tag annotation). Tag counting and tag annotation is combined to obtain the gene expression profile.

for the full protocol. Therefore, SAGE is not suitable for studying many samples at a time.

## SuperSAGE

Although SAGE doubtless is a useful technique for transcriptomics, it nevertheless owns a shortcoming. The size of the SAGE tag, 15 bp, is frequently too short to unequivocally identify the gene of origin. This tag size does not cause much problem in organisms with lower gene numbers (e.g. yeast), but does in organisms with more complex genomes. If entered for a BLAST search as a query, the same tag frequently matches two or more gene sequences, which confounds further analysis. If SAGE is applied to organisms, for which no DNA database is available, it is necessary to recover a longer DNA sequence adjacent to the tag by experiment, and to further annotate this longer fragment by BLAST search. Previously, several polymerase chain reaction (PCR) techniques were reported to recover cDNA fragment from 15 bp SAGE tags (van den Berg *et al.*, 1999; Chen *et al.*, 2000). However, it appeared always difficult to determine the appropriate conditions for a specific amplification of cDNAs from each gene because the SAGE tag primers were too short. The 15 bp DNA fragment is not suitable as an oligonucleotide probe for screening a cDNA library either. To improve the obvious disadvantage of conventional SAGE, Saha *et al.* (2002) replaced the tagging enzyme BsmFI with another Type IIS enzyme, MmeI, to isolate 20 bp tags. This modified version of SAGE, dubbed LongSAGE, was a steplet forward. However, digestion of a DNA fragment with MmeI generates a two-nucleotide recessed 5′ terminus, which is difficult to fill in. Now that SAGE principle requires blunting of the 5′-termini of the tags, which, however, reduces the LongSAGE tag size to 18 bp. In practice, LongSAGE is performed without blunting the 5′ ends of the tags, allowing the isolation of 20 bp tag at the cost of faithful gene expression profiling. Therefore, the resulting 20 bp tag information is only used to help annotate the 15 bp tag obtained by the original SAGE procedure.

To circumvent the problems of SAGE and LongSAGE, we have developed a method named 'SuperSAGE' (Matsumura *et al.*, 2003b). In SuperSAGE, the Type III restriction endonuclease EcoP15I (Meisel *et al.*, 1992; Mücke *et al.*, 2001) is used as the tagging enzyme. Among the restriction enzymes, EcoP15I has the longest distance so far reported between the recognition and cleavage sites. The use of EcoP15I allowed us to isolate 26-bp tags from transcripts (Fig. 1). EcoP15I digestion generates a two-nucleotide recession at the 3′ terminus, which is easy to fill in. Therefore, with SuperSAGE, we can obtain faithful gene expression profiles based on the 26 bp tags. We applied SuperSAGE to profile rice gene expression. Rice is the first crop species for which the whole genomic sequence became available (Goff *et al.*, 2002; Yu *et al.*, 2002). To demonstrate that the 26 bp tag size of Super-SAGE allows a highly reliable tag-to-gene annotation, the following *in silico* experiment was performed. Fifty Super-SAGE tags (26 bp) were randomly selected from rice SuperSAGE data (Matsumura *et al.*, 2003b). These DNA sequences were truncated from the 3′ ends so that the tag sizes were 20, 18 and 15 bp long respectively (Table 1). A tag size of 15 bp corresponds to that produced in the conventional SAGE protocol. The 18 or 20 bp tags are equivalent to LongSAGE tags, when the linker-tag fragments were ligated to each other with or without end-blunting respectively. The tags of different sizes were BLASTed against the entire body of DNA sequences deposited in GenBank representing DNA sequences from more than 130 000 species. The number of species containing DNA sequences with a perfect match to a tag of a given size were counted, and the average and maximum numbers of species were obtained across the 50 tag sequences. The conventional SAGE tag (15 bp) matched DNA sequences of 7.5 species on average, with a maximum of 23 species. All of the 50 tags corresponded to two or more species (Table 1). The 18 bp tags matched 2.4 species on average, with a maximum of 14 species. Twenty-six tags out of 50 corresponded to two or more species. The 20 bp tags matched 1.3 species on average, with a maximum of four species. Only seven tags out of 50 still corresponded to more than two species, indicating a great improvement over the original SAGE tag length (15 bp). However, note that LongSAGE with 20 bp tag does not necessarily produce accurate gene expression profiles (see above). The 26 bp tags of our SuperSAGE method matched 1.1 species on average, with a maximum of only two species. As few as three tags out of 50 corresponded to the DNA sequences of more than two species. These results clearly show that the information content in

**Table 1.** Summary of BLAST search of 50 rice SAGE tags for the entire body of GenBank data.

|  | Tag size (bp) | | | |
| --- | --- | --- | --- | --- |
|  | 15 | 18 | 20 | 26 |
| Average number of species with DNA sequence perfectly matching the tag | 7.5 | 2.4 | 1.3 | 1.1 |
| Maximum number of species with DNA sequence perfectly matching the tag | 23 | 14 | 4 | 2 |
| Number of tags for which DNA sequences of more than two species showed perfect matches | 50 | 26 | 7 | 3 |

the 26 bp DNA tag sequence greatly improves the efficiency of gene annotation of the tags. The 26 bp tags matched DNA sequences of only one species (which in fact was rice, *Oryza sativa*) on average, and in most cases identified a single gene of the particular species. Thus, the annotation of the tag sequence can be carried out almost perfectly. Tag annotation in SuperSAGE can be performed against EST sequence database as well as against whole genome sequences.

## Simultaneous monitoring of host–pathogen gene expression by SuperSAGE

The high information content of the 26 bp SuperSAGE tag allows the simultaneous gene expression analysis of two or more eukaryotic organisms. As an example, we applied SuperSAGE to study gene expression profiles of both rice plants infected with blast disease and the causative fungus *Magnaporthe grisea* (Matsumura *et al.*, 2003b). A draft of the whole genome sequence of *M. grisea* is also available (Martin *et al.*, 2002). After isolating a total of 12 119 tags from blast-infected rice leaves, each tag was annotated by BLAST search for all the genome sequences of rice and *M. grisea*. As expected, the majority of the tags were annotated to rice genes, while 74 tags did not match rice but *M. grisea* sequences. By increasing the number (e.g. 10-fold) of total tags to be analysed in a future study, we will be able to address the details of rice – *M. grisea* interaction transcriptome.

   The first contact of pathogen and host occurs in a limited number of host cells. It is important therefore to obtain gene expression profiles of host and pathogen at such a front line of the battle ensuing any infection. Taking the example of rice blast disease, we will explain in some detail what could be done with SuperSAGE (Fig. 2). Conidial spores of *M. grisea*, an ascomycete fungus, germinate on rice leaves. At the end of the so-called germination tube, an oval structure called appressorium develops. Extremely high turgor pressure inside the appressorium drives the drilling of the appearing so-called infection hyphae through the plant's rigid cell wall, a prerequisite for the colonization of the host by the fungus. Two types of interaction between *M. grisea* and rice exist. In the *compatible interaction*, fungal mycelia keep developing inside host cells, and infection is established. In this case, the host is susceptible and the pathogen is defined as virulent. On the other hand, in the *incompatible interaction*, host cells perceive invasion of the pathogen by some means, and rapid host cell death around the attempted infection site is induced to confine the pathogen. Here, the host is resistant and the pathogen is called avirulent. A careful and in-depth study of the changes in the transcriptional activity of host and pathogen genes in compatible and incompatible interactions,
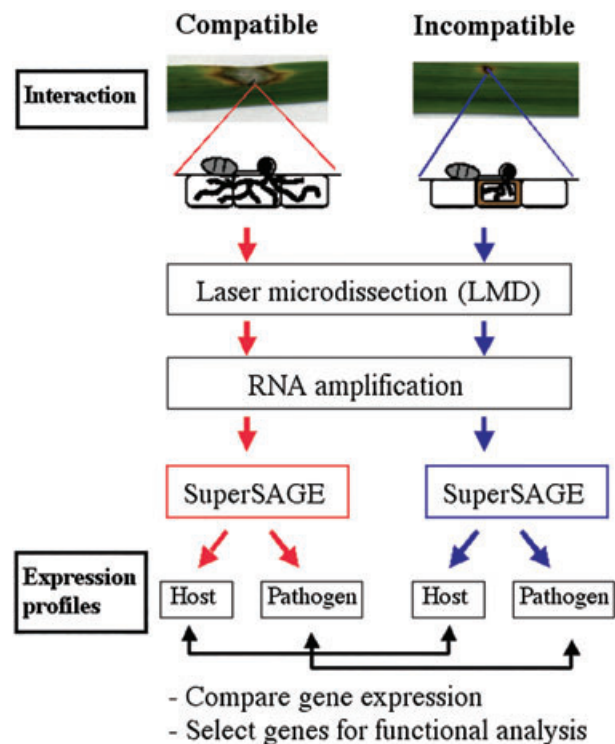


**Fig. 2.** Comparison of gene expression profiles of compatible and incompatible interactions between host and pathogen. A few host (rice) cells interacting with the fungal pathogen (*Magnaporthe grisea*) are isolated by laser microdissection (LMD). Messenger RNAs obtained from the cells are amplified, and subjected to SuperSAGE. SuperSAGE allows separation of host and pathogen transcripts, and identifies a subset of host and pathogen genes that are specifically up- or downregulated in compatible or incompatible interactions respectively.

respectively, is necessary to understand the molecular mechanisms underlying these interactions. It is now possible to isolate even a single cell by laser microdissection (LMD, Emmert-Buck *et al.*, 1996; Nakazono *et al.*, 2003). RNA extracted from LMD-isolated tissue could be amplified five to six orders of magnitudes with high fidelity by T7 RNA polymerase-amplified (a)RNA amplification protocol (Wang *et al.*, 2000), which could be subjected to SAGE (Vilain *et al.*, 2003). Thus, starting from 5 ng of total RNA extracted from tissue collected by LMD, we expect 3 μg of aRNA suitable for SuperSAGE. SuperSAGE would allow separation of rice and *M. grisea* transcripts. By conducting SuperSAGE experiments for compatible and incompatible interactions, we would be able to compare gene expression profiles of susceptible and resistant rice cells as well as virulent and avirulent fungi. By doing so, we should be able to identify host and pathogen genes that are specifically up- and downregulated in compatible and incompatible interactions respectively. From among these identified genes, we can select those that we want to functionally analyse in detail. The

very same strategy is applicable to any interactions between eukaryotic organisms and between eukaryotic host and viruses (e.g. permissive and non-permissive interactions).

A substantial part of mammalian infectious diseases is caused by bacteria. Although some bacterial mRNA have poly(A) sequences at the 3′ end, these sequences are only short lived and used as signal for mRNA degradation. Therefore, it is difficult to monitor bacterial gene expression with the SAGE protocol in which the 3′ end of mRNA is collected on the basis of poly(A) tails. Technical improvement is needed in the simultaneous analysis of gene expression of eukaryotic and prokaryotic organisms.

## Use of SuperSAGE for organisms without DNA database

In organisms for which no DNA database is yet available, the 26 bp SuperSAGE tag can immediately be used as a specific 3′-RACE (Rapid amplification of cDNA ends) primer to rapidly recover a longer cDNA fragment. This cDNA fragment, in turn, could serve as query for a BLAST search to identify the gene of origin. Using the Super-SAGE tag as the 3′-RACE primer, we successfully carried out gene expression profiling in *Nicotiana benthamiana*, a non-model organism without extensive DNA data sets that had been treated with the protein elicitor INF1 from *Phytophthora infestans* (Matsumura *et al.*, 2003b).

## Possible application of SuperSAGE

As mentioned above, the major drawback of SAGE as compared with microarrays is that SAGE can be adapted to only few samples at a time. However, we sometimes wish to compare gene expression profiles of multiple samples. For such purposes, it is possible to design a microarray spotted with cDNA fragments or oligonucleotides that were selected on the basis of SuperSAGE information. In the case of host–pathogen interaction, for instance, we perform SuperSAGE at a given time after the inoculation of the pathogen. By comparing the obtained profile with that of an appropriate control (e.g. mock-inoculated), we can identify SuperSAGE tags that are differentially represented in the two samples. cDNA fragments of the genes corresponding to these subset of tags could be PCR-amplified and spotted onto the array. Otherwise, 26 bp oligonucleotides corresponding to the tag sequences could directly be immobilized on the array. By using these arrays, we would be able to study the kinetics of gene expression during host–pathogen interaction(s). By filtering with SuperSAGE, we can reduce the number of genes subjected to array analysis from, for instance, ~30 000 for human (Crollius *et al.*, 2000) or 30 000–50 000 for rice

(Goff *et al.*, 2002) to ~1000. This approach seems especially helpful in non-model organisms for which infrastructure such as DNA database and cDNA clones are not available.

Gene expression profiling gives only a hint how to select genes for further functional analysis. After target genes are selected, their function is usually tested by reverse-genetics. At the moment the most high-through-put method of reverse genetics is RNAi or gene knockdown. It is known that RNAi is mediated by 21–23 nucleotide long double-stranded (ds) RNA (short interference RNA; siRNA), and administration of 20–28 mer dsRNA to cells can trigger RNAi in *Caenorhabditis elegans* (Kamath *et al.*, 2003), *Drosophila melanogaster* (Lum *et al.*, 2003) and human (Paddison *et al.*, 2002). In this regard, the size of a SuperSAGE tag (26 bp) is relevant. It is long enough (> 21–23 bp) to cause specific RNAi, so that SuperSAGE tag sequences could directly be used for the synthesis of dsRNA to be delivered to the cells to knockdown the gene corresponding to the tag (Fig. 3). SuperSAGE analysis of host–pathogen interac-
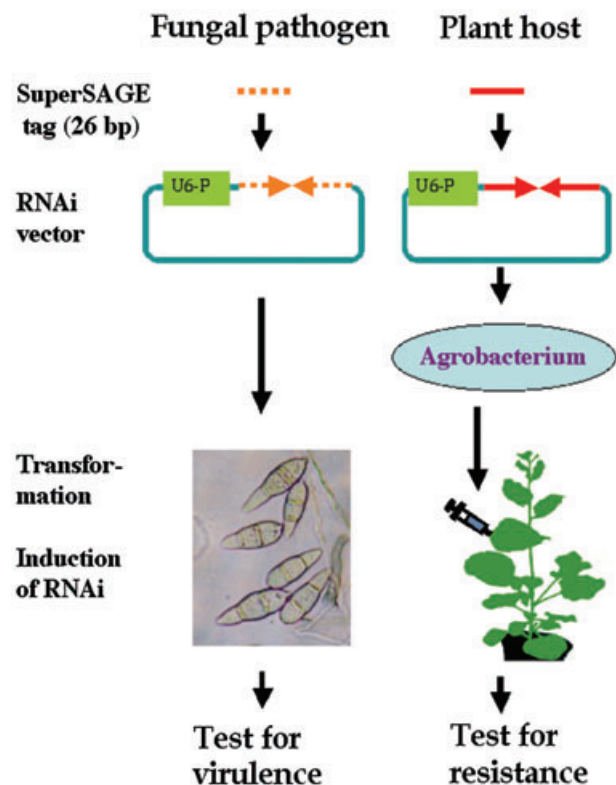


**Fig. 3.** SuperSAGE-RNAi. SuperSAGE tags are long enough to be directly used for synthesizing short-hairpin (sh) RNA, which is expressed in eukaryotic cells to induce RNA interference with the corresponding gene. Tag sequence is cloned into expression vector and transferred directly into fungal cell (left) or transferred to plant cells via *Agrobacterium* (right). After the RNAi of the corresponding gene is established, virulence (fungus) and resistance (plant) can be tested.

tion(s) would identify 26 bp tags, which interest researchers by their expression information. Oligonucleotides corresponding to the tags are synthesized and cloned into an appropriate vector (knockdown vector), so that short-hairpin (sh) RNA could be expressed in the cells by a strong promoter like the U6 snRNA promoter. After transformation of host or eukaryotic pathogen with the knockdown vector, their disease resistance (host) or virulence (pathogen) could be tested within a short time. We have coined this direct use of SuperSAGE tag sequences for RNAi 'SuperSAGE-RNAi'. Furthermore, use of SuperSAGE tags for RNAi would be most useful in the 'large-scale RNA-interference-based screen' (Berns *et al.*, 2004; Paddison *et al.*, 2004). In this screening strategy, a library made in a mammalian expression vector harbouring a total of 5000–10 000 different shRNA is generated, and this library is transfected into mammalian cells en masse. After selection with a selectable marker, only the cells carrying the vector can survive. Each surviving cell should express at least one of 5000–10 000 shRNAs, whereby the gene corresponding to the particular shRNA is knocked down. After that, an appropriate regime is applied so that wild-type cells can no longer survive. Cells surviving under such screen are those expressing particular shRNAs derived from genes necessary for cell death in the screening scheme. Recovery of the vector from surviving cells followed by its sequencing shows the shRNA sequence. From this, we can infer which gene is necessary for cell death in the screen. SuperSAGE would effectively allow to select the shRNA population for library construction. For instance, we treat cells with a particular screening regime and perform SuperSAGE to identify the tags that are over- or under-represented in the cells under the treatment. Oligonucleotides corresponding to such SuperSAGE tags are synthesized, and cloned into the vector to establish the library for the 'RNA-interference-based screen'.

## Concluding remarks

SuperSAGE is a modification of the conventional SAGE procedure, whereby the tag size of 15 bp of the latter is increased to 26 bp owing to the Type III endonuclease EcoP15I as the tagging enzyme. Yet, this increase of tag size by 11 bp has a tremendous impact on the utility and versatility of the SAGE technology, most notably for the analysis of host–pathogen (or host–parasite) interactions. The information content of the 26 bp SuperSAGE tag is sufficient to allow simultaneous gene expression analysis of a host and its pathogen. Also, a SuperSAGE tag is long enough to cause RNAi in eukaryotic cells. We predict that this technology will be an important bridge between transcriptome and gene functional analysis of host–pathogen interactions.

## References

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

Bachem, C.W., van der Hoeven, R.S., de Bruijn, S.M., Vreugdenhil, D., Zabeau, M., and Visser, R.G. (1996) Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant J* **9:** 745–753.

van den Berg, A., Visser, L., and Poppema, S. (1999) Serial analysis of gene expression: rapid RT-PCR analysis of unknown SAGE tags. *Nucleic Acids Res* **27:** e17.

Bergelson, J., Kreitman, M., Stahl, E.A., and Tian, D. (2001) Evolutionary dynamics of plant R-genes. *Science* **292:** 2281–2284.

Berns, K., Hijmans, E.M., Mullenders, J., Brummelkamp, T.R., Velds, A., Heimerikx, M., *et al.* (2004) A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428:** 431–437.

Birch, P.R.J., and Kamoun, S. (2000) Studying interaction transcriptomes: coordinated analyses of gene expression during plant–microorganism interactions. In *New Technologies for Life Sciences: A Trends Guide* (supplement to Elsevier Trends Journals, December 2000) London: Elsevier, pp. 77–82.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., *et al.* (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* **18:** 630–634.

Bryant, P.A., Venter, D., Robins-Browne, R., and Curtis, N. (2004) Chips with everything: DNA microarrays in infectious diseases. *Lancet Infect Dis* **4:** 100–111.

Chen, J.J., Rowley, J.D., and Wang, S.M. (2000) Generation of longer cDNA fragments from serial analysis of gene expression tags for gene identification. *Proc Natl Acad Sci USA* **97:** 349–353.

Crollius, H.R., Jaillon, O., Dasilva, S., Ozouf-Costaz, C., Fizames, C., Fischer, C., *et al.* (2000) Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*. *Genome Res* **10:** 939–949.

Cummings, C.A., and Relman, D.A. (2000) Using DNA microarrays to study host–microbe interactions. *Emerg Infect Dis* **6:** 513–525.

Cummings, L.M., and Wirth, D.F. (2001) Serial analysis of gene expression (SAGE) in *Plasmodium falciparum*: application of the technique to A-T rich genomes. *Mol Biochem Parasitol* **113:** 23–34.

Diehn, M., and Relman, D.A. (2001) Comparing functional genomic datasets: lessons from DNA microarray analyses of host–pathogen interactions. *Curr Opin Microbiol* **4:** 95–101.

Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., *et al.* (1996) Laser capture microdissection. *Science* **274:** 998–1001.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. Japonica). *Science* **296:** 92–100.

Irie, T., Matsumura, H., Terauchi, R., and Saitoh, H. (2003) Serial analysis of gene expression (SAGE) of *Magnaporthe grisea*: genes involved in appressorium formation. *Mol Genet Genomics* **270:** 181–189.

ISPP (1998) International Society for Plant Pathology task force on global food security [WWW document]. URL http://www.isppweb.org/foodsecurity_background.asp

Kahn, R.A., Fu, H., and Roy, C.R. (2002) Cellular hijacking: a common strategy for microbial infection. *Trends Biochem Sci* **27:** 308–314.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., *et al.* (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421:** 231–237.

Kamoun, S., Hraber, P., Sobral. B., Nuss, D., and Govers, F. (1999) Initial assessment of gene diversity for the oomycete pathogen *Phytophthora infestans* based on expressed sequences. *Fungal Genet Biol* **28:** 94–106.

Kato-Maeda, M., Gao, Q., and Small, P.M. (2001) Microarray analysis of pathogens and their interaction with hosts. *Cell Microbiol* **3:** 713–719.

Liang, P., and Pardee, A.B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257:** 967–971.

Lum, L., Yao, S., Mozer, B., Rovescalli, A., Von Kessler, D., Nirenberg, M., and Beachy, P.A. (2003) Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299:** 2039–2045.

Martin, S.L., Blackman, B.P., Rajagopalan, R., Houfek, T.D., Sceeles, R.G., Denn, S.O., *et al.* (2002) MagnaportheDB: a federated solution for integrating physical and genetic map data with BAC and derived sequences for the rice blast fungus *Magnaporthe grisea*. *Nucleic Acids Res* **30:** 121–124.

Matsumura, H., Nirasawa, S., and Terauchi, R. (1999) Technical advance: transcript profiling of rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression. *Plant J* **20:** 719–726.

Matsumura, H., Nirasawa, S., Kiba, A., Urasaki, N., Saitoh, H., Ito, M., *et al.* (2003a) Overexpression of Bax inhibitor suppresses the fungal elicitor-induced cell death in rice (*Oryza sativa* L.) cells. *Plant J* **33:** 425–434.

Matsumura, H., Reich, S., Ito, A., Saitoh, H., Kamoun, S., Winter, P., *et al.* (2003b) Gene expression analysis of plant host–pathogen interactions by SuperSAGE. *Proc Natl Acad Sci USA* **100:** 15718–15723.

Meisel, A., Bickle, T.A., Krüger, D.H., and Schroeder, C. (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature* **355:** 467–469.

Mücke, M., Reich, S., Moncke-Buchner, E., Reuter, M., and Krüger, D.H. (2001) DNA cleavage by type III restriction-modification enzyme EcoP15I is independent of spacer distance between two head to head oriented recognition sites. *J Mol Biol* **312:** 687–698.

Nakazono, M., Qiu, F., Borsuk, L., and Schnable, P.S.

(2003) Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize. *Plant Cell* **15:** 583–596.

Paddison, P.J., Caudy, A.A., and Hannon, G.J. (2002) Stable suppression of gene expression by RNAi in mammalian cells. *Proc Natl Acad Sci USA* **99:** 1443–1448.

Paddison, P.J., Silva, J.M., Conklin, D.S., Schlabach, M., Li, M., Aruleba, S., *et al.* (2004) A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428:** 427–431.

Patankar, S., Munasinghe, A., Shoaibi, A., Cummings, L.M., and Wirth, D.F. (2001) Serial analysis of gene expression in *Plasmodium falciparum* reveals the global expression profile of erythrocytic stages and the presence of antisense transcripts in the malarial parasite. *Mol Biol Cell* **12:** 3114–3125.

Patino, W.D., Mian, O.Y., and Hwang, P.M. (2002) Serial analysis of gene expression: technical consideration and applications to cardiovascular biology. *Circ Res* **91:** 565–569.

Polyak, K., and Riggins, G. (2001) Gene discovery using the serial analysis of gene expression technique: implications for cancer research. *J Clin Oncol* **19:** 2948–2958.

Polyak, K., Xia, Y., Zweier, J.L., Kinzler, K.W., and Vogelstein, B. (1997) A model for p53 induced apoptosis. *Nature* **389:** 300–305.

Saha, S., Sparks, A.B., Rago, C., Akmaev, V., Wang, C.J., Vogelstein, B., *et al.* (2002) Using the transcriptome to annotate the genome. *Nat Biotechnol* **20:** 508–512.

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

Steen, B.R., Lian, T., Zuyderduyn, S., MacDonald, W.K., Marra, M., Jones, S.J., *et al.* (2002) Temperature-regulated transcription in the pathogenic fungus *Cryptococcus neoformans*. *Genome Res* **12:** 1386–1400.

Thomas, S.W., Glaring, M.A., Rasmussen, S.W., Kinane, J.T., and Oliver, R.P. (2002) Transcript profiling in the barley mildew pathogen *Blumeria graminis* by serial analysis of gene expression (SAGE). *Mol Plant Microbe Interact* **15:** 847–856.

Velculescu, V., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995) Seraial analysis of gene expression. *Science* **270:** 484–487.

Velculescu, V., Zhang, L., Zhou, W., Vogelstein, B., Basrai, M.A., Basette, D.E., *et al.* (1997) Characterization of the yeast transcriptome. *Cell* **88:** 243–251.

Vilain, C., Libert, D., Venet, D., Costagliola, S., and Vassart, G. (2003) Small amplified RNA-SAGE: an alternative approach to study transcriptome from limiting amount of mRNA. *Nucleic Acids Res* **31:** e24.

Wan, J., Dunning, F.M., and Bent, A.F. (2002) Probing plant–pathogen interactions and downstream defense signaling using DNA microarrays. *Funct Integr Genomics* **2:** 259–273.

Wang, E., Miller, L.D., Ohnmacht, G.A., Liu, E.T., and Marincola, F.M. (2000) High-fidelity mRNA amplification for gene profiling. *Nat Biotechnol* **18:** 457–459.

WHO (2003) *Communicable Diseases Cluster: Highlights of Communicable Disease Activities, Major Recent Achievements*. Geneva, Switzerland: World Health Organization.

Woolhouse, M.E.J., Webster, J.P., domingo, E., Charlesworth, B., and Levin, B.R. (2002) Biological and biomedical implications of the co-evolution of pathogens and their host. *Nature Genet* **32:** 569–577.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *Indica*). *Science* **296:** 79–92.

Zhang, X., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., *et al.* (1997) Gene expression profiles in normal and cancer cells. *Science* **276:** 1268–1271.