RESULTS AND DISCUSSION

Background

This work was performed within the network of two projects: (i) a GABI-MASC ("Genomanalyse im biologischen System Pflanze - Max-Planck Gesellschaft Arabidopsis Consortium"), focused at creating a genotyping instrument for the plant *Arabidopsis thaliana* and (ii) a joint German-Russian program, aimed at revealing the spectrum and frequencies of alleles of clinically important SNPs in the populations of Eurasia. In both projects our group was in charge of developing a convenient and cost-effective technology for SNP detection and providing the other participants with ready-to-use genotyping kits.

The genotyping method had to be adequate for analysis of a moderate number (tens-hundreds) of SNPs in a large number of DNA samples and was supposed to be used by molecular biology laboratories, for which SNP genotyping is not the only type of work.

Table 2. Genotyping approaches examined in the study.

Genotyping approach	Limitations
arrested primer extension on microarray	- complex target preparation (PCR amplification and removal of one DNA strand) - hybridization of several targets to immobilized oligonucleotides results in high background
allele-specific primer extension on genomic DNA and detection of extended products by hybridization with microarray	- unacceptably high cross-hybridization due to multiple duplications in *A. thaliana* genome - sensitivity too low for genotyping of human DNA
allele-specific PCR	- individual assay optimization required for each locus - specificity varies for different loci - limited multiplexing possibilities
allele-specific ligation of linear probes on genomic DNA with subsequent amplification and detection of ligated product by hybridization with microarray	- complex multistep procedure - contamination risk during microarray processing
allele-specific ligation of circular probes (padlocks) on genomic DNA with subsequent amplification, combined with TaqMan detection	- padlocks give higher background than linear probes - padlocks are expensive and difficult to prepare

The most critical demands of our collaborators concerned (i) reliability of the method and (ii) its sensitivity (minimal amount of DNA required for analysis).

Several genotyping approaches were examined (for a list, see Table 2). Although all of them enable genotyping of SNPs, some insurmountable limitations made them inappropriate for routine laboratory practice.

We have finally developed the ligation detection reaction (LDR) - TaqMan SNP genotyping method. On the basis of this technique, two genotyping kits (a whole-genome kit for *A. thaliana* and a kit for 9 human loci) were prepared, tested, and transferred to our collaborators.

LDR-TaqMan SNP detection method

The scheme of the LDR-TaqMan method is presented in Figures 13 and 14, showing the "one tube – one locus" and the "one tube – many loci" formats, respectively (detailed descriptions are below). In both cases, the first step of the procedure is a ligation detection reaction (LDR) converting the biallelic state of each SNP locus into the bimarker state of ligated detector oligonucleotides (DOs). The state of the markers is determined by a TaqMan assay during subsequent PCR amplification of ligated DOs.

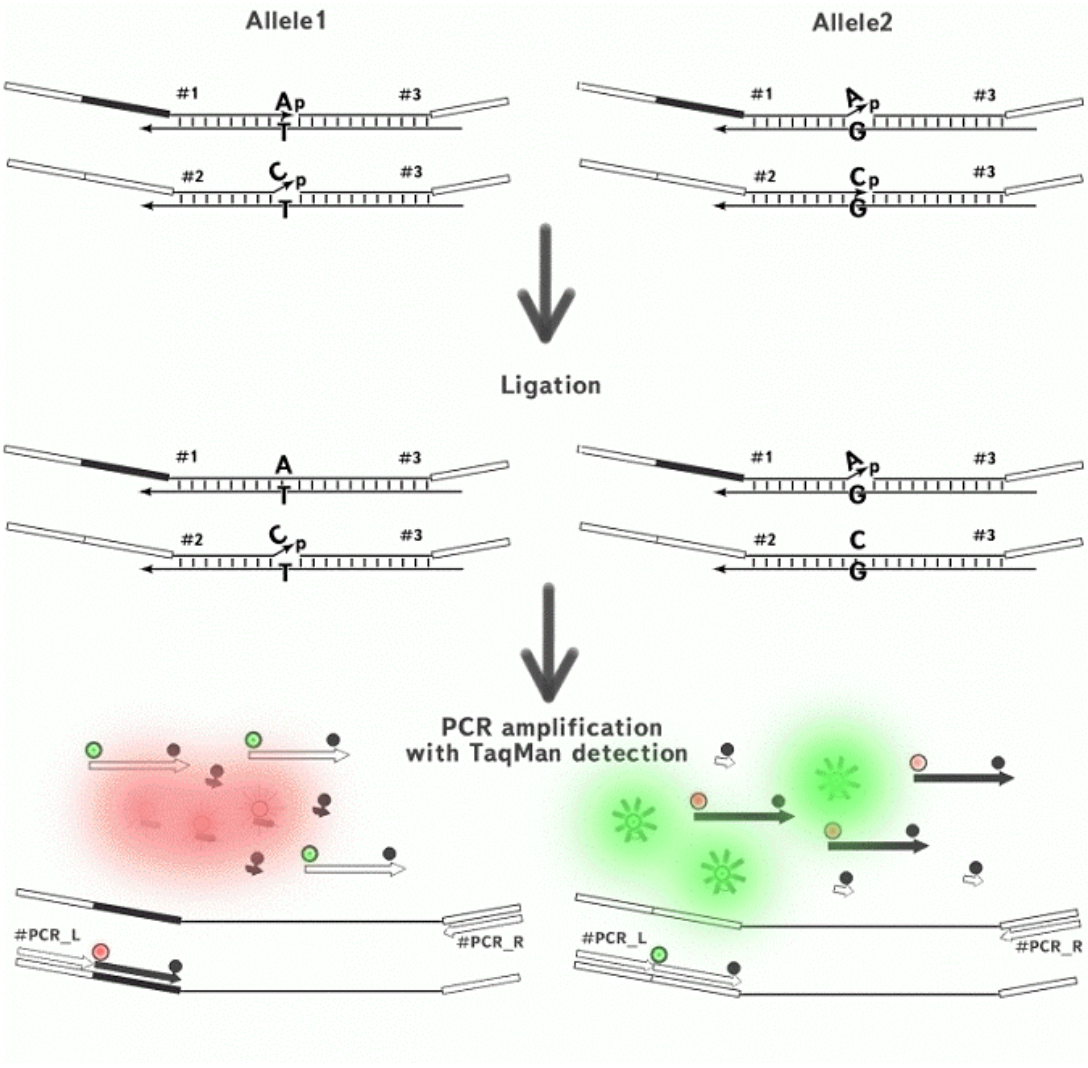

Figure 13. Scheme of the LDR-TaqMan SNP detection method: "one tube – one locus" variant.

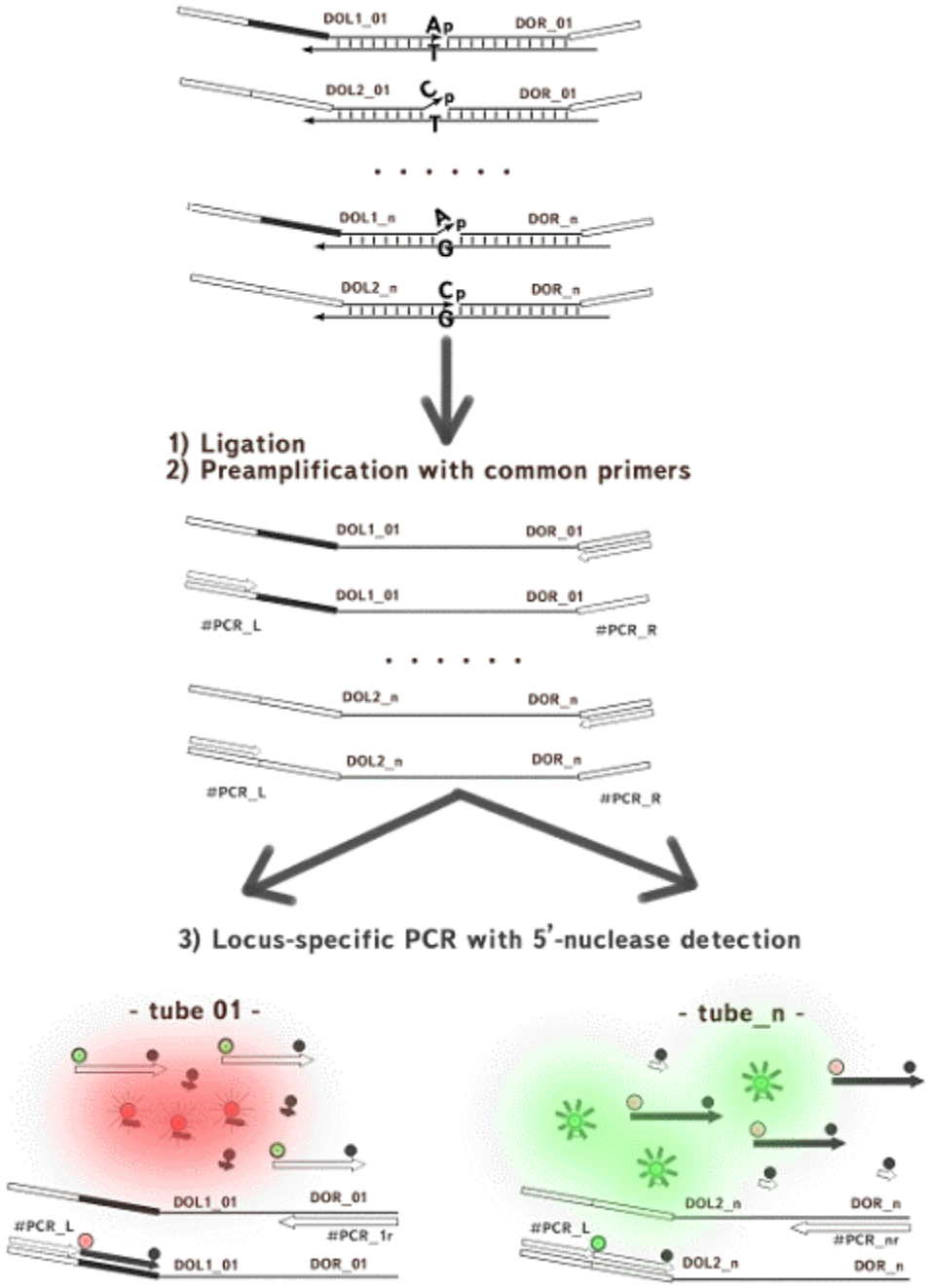

Figure 14. Scheme of the LDR-TaqMan SNP detection method: "one tube – many loci" variant.

The structure of detector oligonucleotides

The same DOs are used both for single and multiplex ligations. Three DOs are required for each biallelic locus: a common "right" DO (DOR) and two different "left" DOs (DOL1 and DOL2) (Figure 15B,C).

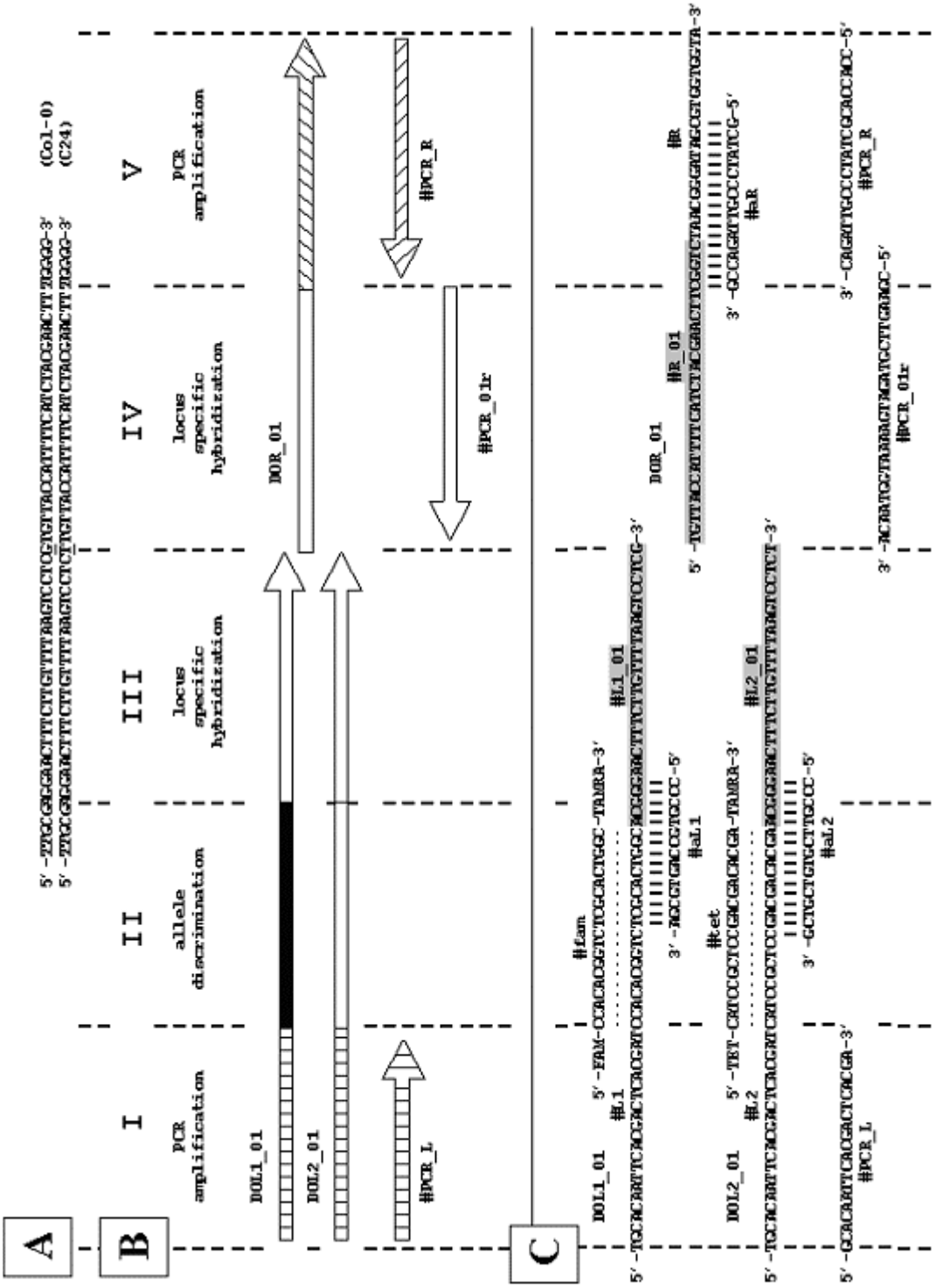

Figure 15. Detector oligonucleotides and PCR primers for *A. thaliana* locus "01" aligned relative to the genomic sequence. **A.** Genomic sequences in the vicinity of the SNP position for *A. thaliana* accessions Col-0 and C24. **B.** Block structure of DOs. Roman numerals indicate structural regions of DOs. **C.** Nucleotide sequences of oligonucleotides. Locus-specific oligonucleotides used for preparation of DOs (#L1_01, #L2_01, #R_01) are marked by a gray background. #L1, #L2 and #R are common primers. #aL1, #aL2 and #aR are adaptors used for ligation-based synthesis of DOs. #fam and #tet are universal TaqMan probes. #PCR_L and #PCR_R are common PCR primers. #PCR_01r is a locus-discrimination PCR primer for the "one tube – many loci" scheme.

The only locus-specific parts of DOs are Regions III and IV, which target DOs to a particular SNP position. Region III is allele-specific: two left DOs have different 3'-terminal nucleotides corresponding to the SNP position. The algorithm for selection of locus-specific parts of DOs is described in "Materials and methods".

The other parts of DOs (Regions I, II and V) are the same for different DO-sets. They are used for the TaqMan assay. The sequence of Region I is identical to the PCR

primer #PCR_L. The sequence of Region V is complementary to the PCR primer #PCR_R. Region II (shown in black and white in Figure 15B) harbors the targets for the universal TaqMan probes (#fam and #tet in Figure 15C).

The "one tube – one locus" procedure (Figure 13)

The "one tube – one locus" procedure is a two-step method (Figure 13) with end point fluorescence detection. The first step is a ligation reaction performed directly on genomic DNA. Depending on the allelic state of the SNP locus, one of the "left" DOs is ligated to the "right" DO forming an amplicon with universal 3'- and 5'-ends. Thus, the task to determine the single nucleotide in the SNP position is substituted by the task to determine which type of Region II (20nt sequence) is present in the amplicon. The second step is the TaqMan assay with universal amplification primers and TaqMan probes. Both steps are performed in one tube: the amplification mixture (*Taq* DNA polymerase, PCR primers #PCR_L and #PCR_R, TaqMan probes #fam and #tet and buffer) is added directly to the ligation reaction. Detection of both alleles in the same reaction helps to eliminate false heterozygote readings [Pickering et al., 2002].

An example of the genotyping results obtained by the "one tube – one locus" procedure is presented in Figure 16 (corresponding fluorescence values are given in Supplement 3). 30 different SNP loci were analyzed on *A. thaliana* genomic DNA. 24 loci ("1"-"14", "17", "18", "20"-"22", "26"-"28", "A" and "B"; see Supplement 1) polymorphic for Columbia (Col-0) and C24 accessions, are evenly distributed along the 1st chromosome. The other 6 loci are located on the 4th chromosome and were used for initial experiments. Loci "C", "D", "E", "F" are polymorphic for Col-0 and C24 and loci "G" and "H" for Col-0 and Landsberg *erecta* (L*er*) accessions. Fam fluorescence signals correspond to Col-0 SNP variants, Tet signals to C24 and L*er* variants.

Genotyping data for all loci affirmed the known genotype. Fluorescence signals for homozygous and heterozygous DNA in Figure 16 are well separated from each other and clearly grouped in three areas. The signals for heterozygous DNA are more dispersed than for homozygous DNA. Independent experiments proved that the location of signals for each particular locus is reproducible. Figure 17 demonstrates reproducibility of results obtained for loci "20" and "22" on different amounts of genomic DNA (5-150ng). Despite a 30-fold difference in starting DNA quantities, signals are tightly grouped.

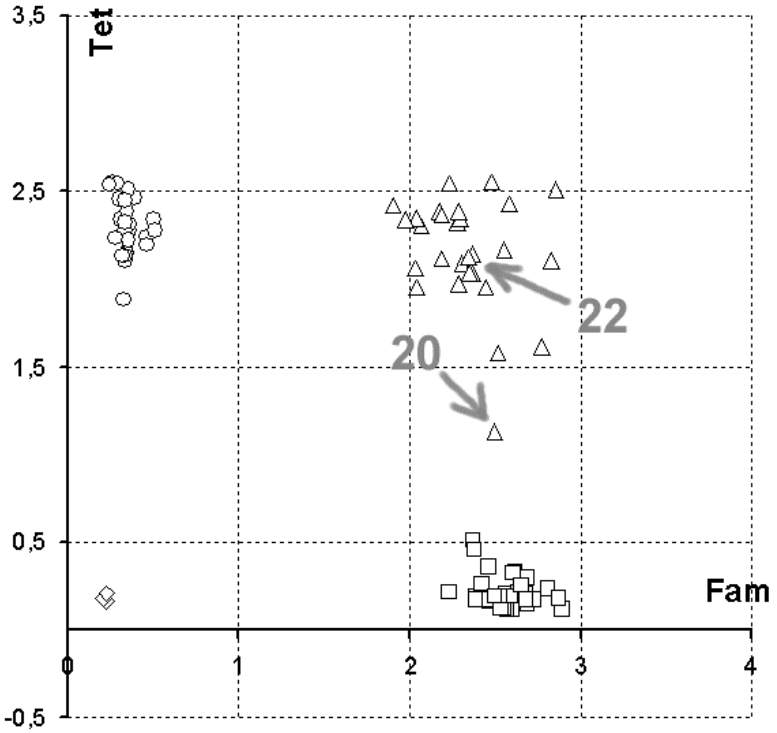

Figure 16. Scatter plot of "one tube – one locus" genotyping results (fluorescence values are given in Supplement 3) for 30 different SNP loci of *A. thaliana*. Genotyping reactions were performed on 50ng of homozygous and heterozygous genomic DNA. Fam and Tet signals are plotted on the X- and Y-axes respectively. Homozygous samples are represented as squares (Col-0) and circles (C24, L*er*), heterozygous (Col-0/C24, Col-0/L*er*) as triangles, and controls without DNA as rhomboids. Arrows show heterozygous signals for SNP loci "20" and "22".

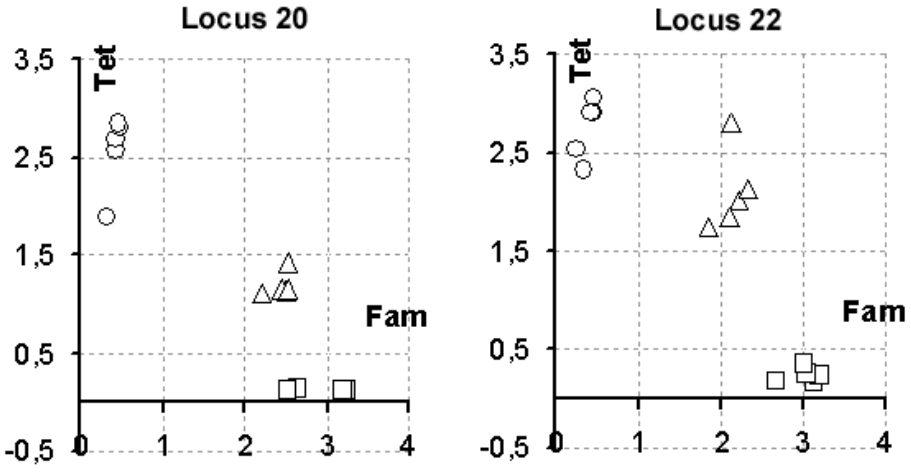

Figure 17. Scatter plots of genotyping data for SNP loci "20" and "22" (fluorescence values are given in Table 3). Genotyping reactions were performed on 5, 17, 50 and 150ng of homozygous and heterozygous genomic DNA. Homozygous samples are represented as squares (Col-0) and circles (C24), heterozygous (Col-0/C24) as triangles.

Heterozygous signals for loci "20" and "22" have different Fam/Tet ratios. It seems that this is caused by unequal amplification of allele-specific amplicons due to sequence difference: they differ in the sequence of Region II and in the nucleotide in the SNP position.

Table 3. Fluorescence values for "one tube – one locus" SNP-genotyping for loci "20" and "22". Results for these loci from Supplement 3 are also included (marked with an asterisk).

Locus	DNA (ng)	Col-0		C24		Col-0l/C24	
		Fam	Tet	Fam	Tet	Fam	Tet
20	150	2.65	0.14	0.43	2.57	2.20	1.12
	50	3.24	0.12	0.43	2.69	2.53	1.42
	50*	2.54	0.13	0.33	1.88	2.50	1.13
	17	3.24	0.12	0.48	2.80	2.46	1.15
	5	3.20	0.13	0.46	2.84	2.53	1.16
22	150	3.14	0.16	0.26	2.53	1.85	1.74
	50	3.20	0.23	0.47	3.06	2.22	2.01
	50*	2.68	0.17	0.34	2.32	2.34	2.12
	17	3.04	0.26	0.46	2.90	2.10	1.84
	5	3.02	0.35	0.44	2.90	2.13	2.80

Other genotyping results obtained by the "one tube – one locus" procedure are given in the "LDR-TaqMan genotyping kits" section. Figure 28 shows the end point signal distribution for 1380 independent genotypings on 10 samples of *A. thaliana* genomic DNA. Figure 29 demonstrates the analogous graph for 1035 independent genotypings on 115 of human DNA samples. In all cases, signals for homozygous and heterozygous DNA are clearly distinguishable.

The "one tube – many loci" procedure (Figure 14)

The overview of the "one-tube - many loci" protocol is shown in Figure 14. The ligation detection reaction (LDR) is performed simultaneously for a number of loci in one tube. The allelic state of each particular locus is determined in a separate TaqMan reaction (step 3 in Figure 14). One of PCR primers is locus-specific, the other PCR primer and TaqMan probes are common for all loci. Locus-specific primers (e.g. #PCR_01r primer in Figure 15C) are used to distinguish particular amplicon within a complex mixture. An intermediate pre-amplification (step 2) is introduced into the

protocol to prevent decrease of sensitivity due to the distribution of the ligation mixture into a number of separate tubes for the locus-specific TaqMan assay.

Pre-amplification of all ligated DOs is carried out with two common primers, which is essential to maintain the relative proportion of different amplicons during PCR. As was shown in Amplification Fragment Length Polymorphism (AFLP) experiments, the initial proportions are maintained for hundreds of amplicons when amplification is performed with common primers [Vos et al., 1995].

If only a few (≤10) loci are multiplexed and/or the amount of DNA is not limited, the pre-amplification step may be omitted [Rickert et al., 2004]. In this case, the procedure consists of two steps and needs less handling and time. The ligation reaction is distributed directly into the tubes with locus-specific amplification mixes. Moreover, shorter "right" DOs (DORs) may be used: Region V is not required and may be excluded.

Results obtained in the parallel genotyping of the 30 SNP loci (the loci are described previously in the "one tube – one locus procedure" section) are shown in Figure 18.

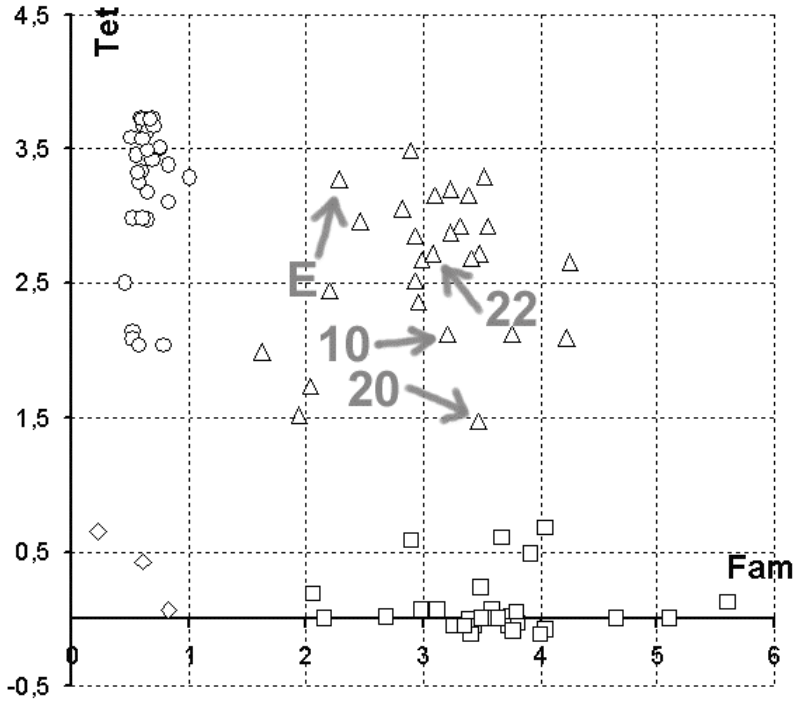

Figure 18. Scatter plot of the "one tube – many loci" genotyping results (fluorescence values are given in Supplement 4). Genotyping of 30 different SNP loci was performed on 50ng of homozygous and heterozygous *A. thaliana* genomic DNA. Homozygous samples are represented as squares (Col-0) and circles (C24), heterozygous (Col-0/C24) as triangles, and controls without ligation (amplification was performed with common PCR primers) as rhomboids. Arrows show heterozygous signals for SNP loci "10", "20", "22", and "E". Signals for locus "A" are not shown.

Signals corresponding to homozygous and heterozygous DNA group into three separate areas.

The dispersion of signals is higher than that for the "one tube – one locus" procedure. Probably mis-priming of locus-specific PCR primers adds to the dispersion caused by unequal amplification of allele-specific amplicons. Again, the location of signals on the graph is reproducible for each locus, as may be seen in Figure 19, where end point fluorescence signals obtained in independent experiments are plotted for loci "10", "20", "22" and "E". Figure 19 also illustrates that the genotype may be reliably determined using as little as 5ng of *A. thaliana* genomic DNA.

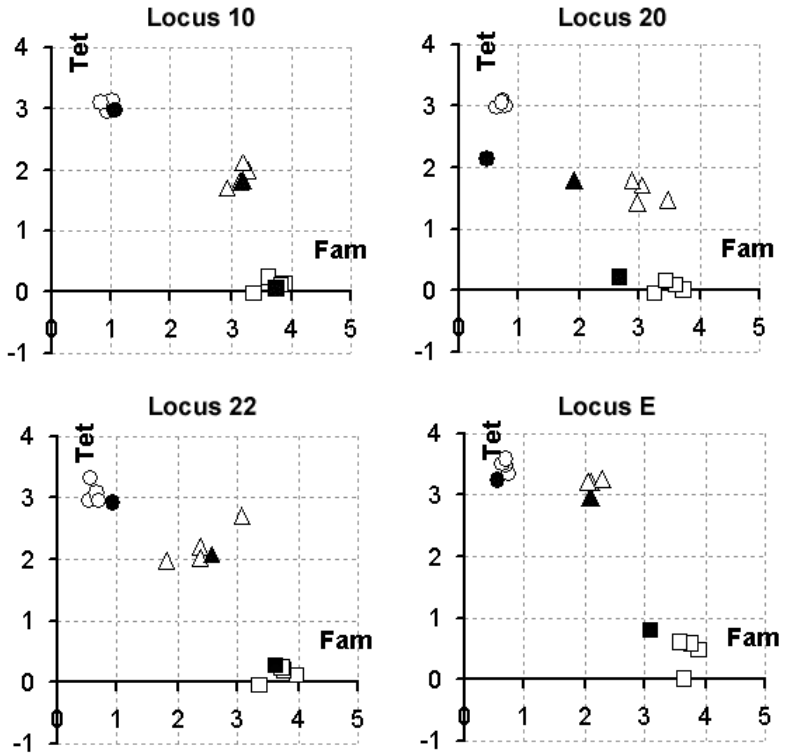

Figure 19. Scatter plots of the "one tube – many loci" genotyping data for SNP loci "10", "20", "22", and "E" (fluorescence values are given in Table 4). Fifteen independent genotyping reactions were performed on 5ng (filled icons), 17ng, 50ng and 150ng of homozygous and heterozygous DNA. Homozygous samples are represented as squares (Col-0) and circles (C24), heterozygous (Col-0/C24) as triangles.

Some DO-sets may be incompatible with others in a multi-locus parallel analysis, since some particular DO may serve as a template for mis-ligation of other DOs. Among 30 SNP loci tested, only one locus (locus "A") showed such a problem. Although the DO-set for locus "A" works very well in the "one tube – one locus" genotyping (Supplement 3), it gives a Fam/Tet ratio of about 4 in the multi-locus

Table 4. Fluorescence values for the "one tube – many loci" SNP-genotyping for SNP loci "10", "20", "22" and "E". Results for these loci from Supplement 4 are also included and marked with an asterisk.

Locus	DNA (ng)	Col-0		C24		Col-0/C24	
		Fam	Tet	Fam	Tet	Fam	Tet
10	150	3.64	0.25	0.94	2.93	2.92	1.70
	50	3.90	0.14	1.02	3.13	3.27	1.99
	50*	3.40	-0.01	0.83	3.10	3.20	2.12
	17	3.85	0.11	1.04	2.98	3.15	1.81
	5	3.75	0.06	1.07	2.96	3.19	1.82
20	150	3.75	0.01	0.76	3.07	3.05	1.70
	50	3.61	0.09	0.77	3.01	2.90	1.80
	50*	3.26	-0.05	0.64	2.97	3.47	1.47
	17	3.46	0.16	0.73	3.06	2.97	1.43
	5	2.68	0.20	0.48	2.12	1.92	1.79
22	150	3.99	0.10	0.54	2.94	1.82	1.98
	50	3.76	0.18	0.67	3.07	2.38	2.20
	50*	3.36	-0.05	0.56	3.32	3.08	2.72
	17	3.75	0.24	0.70	2.95	2.38	2.02
	5	3.63	0.26	0.94	2.92	2.58	2.07
E	150	3.91	0.48	0.76	3.34	2.10	2.96
	50	3.76	0.57	0.70	3.48	2.12	3.21
	50*	3.65	0.01	0.65	3.49	2.29	3.27
	17	3.57	0.61	0.70	3.59	2.07	3.22
	5	2.11	0.80	0.56	3.24	2.09	2.95

reaction on heterozygous DNA (Supplement 4). Problematic DOs may be re-designed (for genotyping on the complementary strand) or corresponding SNPs may be analyzed separately.

The main point of the parallel analysis is that in this format, considerably less genomic DNA is required for genotyping of each particular locus: 17ng of DNA for genotyping of 30 loci means 0.6ng per locus. As these tests were performed well within the sensitivity limits of the experiment, it is conceivable that the amount of DNA could be reduced even further. Performing a reaction in one tube for several loci also reduces handwork and saves consumables.

Discussion of the LDR-TaqMan protocol

The LDR-TaqMan is a novel SNP genotyping method suitable both for individual and multiplex analysis. It exploits the ligation detection reaction for SNP discrimination and the TaqMan assay for visualization. The method has proven to be reliable and

reproducible. It requires no locus-specific optimization, and it is amenable to automation.

In terms of benchwork, the LDR-TaqMan procedure is quick and user-friendly. Genotyping may be performed in 96- or 384-well microtitre plates. A homogeneous real-time or end point fluorescence detection eliminates the need for purification or separation. The "one tube – one locus" procedure includes only two pipetting steps (Figure 20A): (i) addition of ligation mixture to the genomic DNA and (ii) addition of amplification mixture to the ligation reaction. The "one tube – many loci" procedure includes one additional pipetting step to perform pre-amplification (Figure 20B). Ligation and amplification mixtures contain thermostable enzymes and may be stored pre-dispensed at –20°C or +4°C.

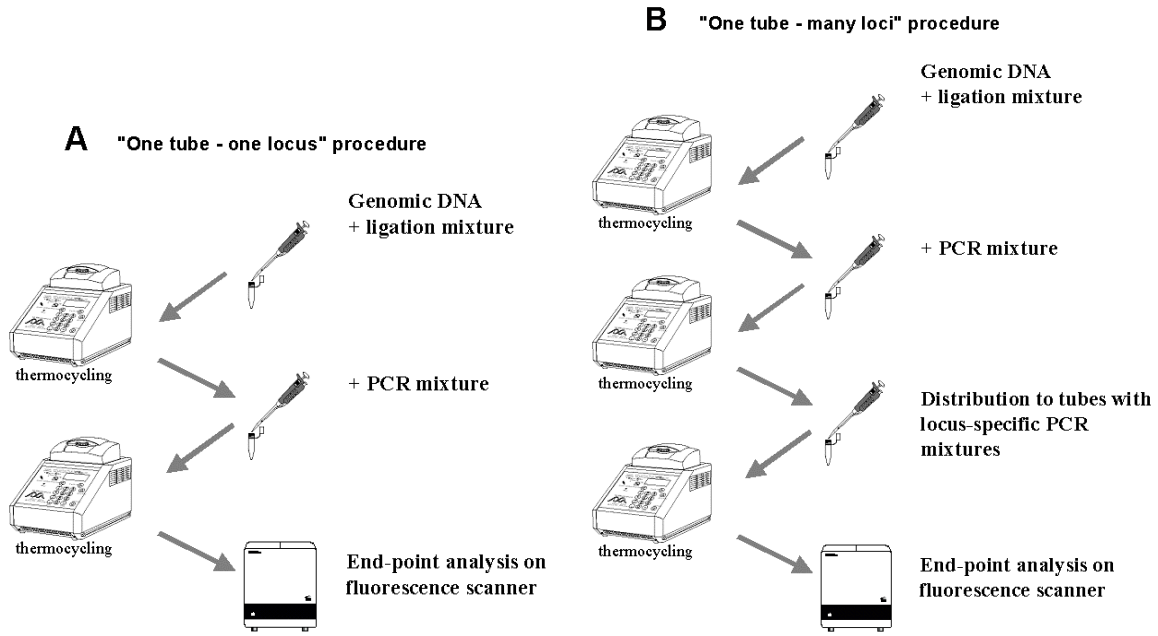

Figure 20. Handling steps of the "one tube – one locus" (A) and the "one tube – many loci"(B) procedures.

The method requires ordinary equipment - a PCR machine and a fluorescence plate reader, which are present or at least affordable in a common laboratory.

SNP discrimination directly on genomic DNA does not involve manipulations with amplified SNP-containing regions. This is crucial for a non-specialized laboratory, since without separation of pre- and post- amplification areas and other special precautions, manipulations with amplified DNA may cause serious contamination problems. The "one tube – one locus" procedure is a closed-tube technique. The "one tube – many loci"

scheme is also contamination-safe: ≤10 cycles of pre-amplification increase the amount of ligated DOs by only approximately 1000-fold.

Figure 21 illustrates the sensitivity of the method: 0.4ng of *A. thaliana* genomic DNA and 2ng of human genomic DNA are enough for genotyping. Sensitivity may be further increased if the ligation detection is performed for several loci on the same aliquot of genomic DNA.

The LDR-TaqMan might be a method of choice for small- and medium-scale projects requiring a moderate level of multiplexing (<1000 loci).

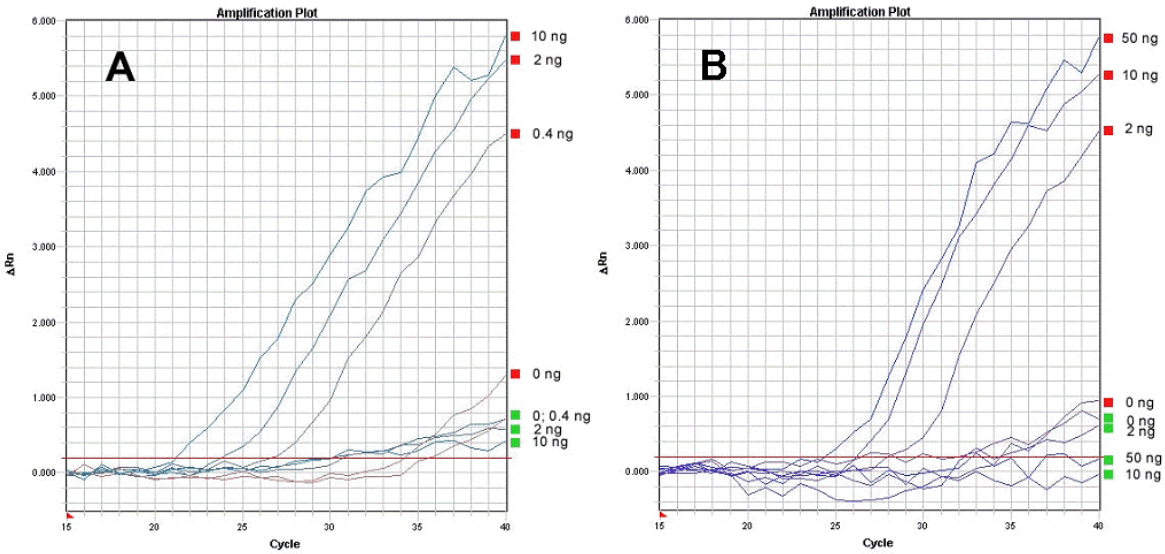

Figure 21. LDR-TaqMan genotyping on different amounts of genomic DNA. Amplification plots show the change of fluorescence versus cycle number. Red squares indicate Fam PCR curves, green indicates Tet (background) curves. (A) *A. thaliana* locus "G" analyzed on 0, 0.4, 2 and 10ng of Col-0 DNA. (B) human locus LCT (–13910 C/T) analyzed on 0, 2, 10 and 50ng of Fam-homozygous DNA.

The LDR-TaqMan method is somewhat similar to the TaqMan SNP assay. The advantages of our approach are: (i) no locus-specific optimization required, (ii) a lower setup cost due to standard PCR and fluorescent primers, and (iii) the possibility of multiplex analysis.

The Amplifluor assay commercially available from Kbioscience also uses universal fluorescent probes for reporting the allelic state of the locus [Myakishev et al., 2001; Hawkins et al., 2002]. This assay is cheaper and more flexible than the "one tube – one locus" LDR-TaqMan, because it (i) is a one-step, one-enzyme procedure, (ii) uses shorter locus-specific oligonucleotides, and (iii) allows for more freedom in the selection of locus-specific PCR primers. On the other hand, the Amplifluor assay requires locus-specific tuning and has no potential for multiplexing [Rickert et al.,

2004]. Our "one tube – many loci" scheme should be more cost-effective than the Amplifluor approach due to the parallel analysis of a number of loci.

In terms of multiplexing and throughput, our method is similar to the recently developed SNPlex assay (Applied Biosystems). SNPlex is also based on LDR for SNP discrimination, followed by PCR amplification of the ligation products with standard primers. The detection of allelic variants is performed by special ZipChute Mobility Modifier probes: after hybridization to the amplified ligation products, hybridized probes are washed away and analyzed by capillary electrophoresis. Allele-specific probes differ in their mobility in a gel and bear a locus-specific fluorophore. The company provides kits capable of multiplexing 48 SNPs in a single reaction.

We have also tried to perform parallel analysis of ligated products: amplified ligated products were labeled, one strand was removed, and products were hybridized to an array of locus-specific probes. The problem is that both gel separation (in SNPlex assay) and hybridization on microarrays (as our early procedure) are complex, multistep procedures, for which the contamination risk must be considered.

The discrimination power of the method depends on the specificity of ligation. We have compared the fidelity (ratio of match/mismatch ligation) of five commercially available thermostable ligases: *Taq* DNA ligase, *Tth* DNA ligase, Ampligase, *Pfu* ligase and Ligase-65. *Pfu* DNA ligase showed the best results (data not shown).

The high fidelity of *Pfu* DNA ligase is confirmed by the fact that it functions in the LDR-TaqMan assay in a remarkably large range of concentrations: 0.004 to 4U of enzyme give the same results both in terms of efficiency and match/mismatch ratio (Figure 22).

Cyclic ligation is performed to increase the sensitivity of the reaction (see "Materials and methods").

Each ligation cycle includes 6 sub-cycles: 65°C for 30 seconds followed by 10 seconds at 74°C. Annealing and ligation of the DOs is performed at 65°C. 74°C is the temperature at which unligated DOs dissociate from the template and ligated DOs remain hybridized. Sub-cycles provide a chance for matched DOs to anneal to the template and be ligated. This increases the yield of the reaction by about two-fold. Another advantage of sub-cycles is that they prevent the increase of background noise. For each DO set, there is an optimal ligation time when the majority of annealed,

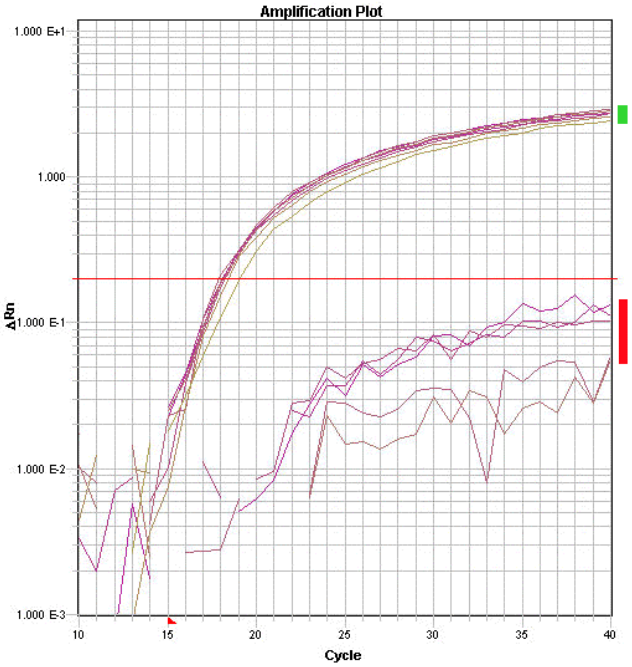

Figure 22. Real-time amplification curves for LDR-TaqMan genotyping of locus "G" on 100ng of C24 DNA. Amplification plot shows the log of the change of fluorescence versus cycle number. Ligation was performed with 4-fold dilutions of *Pfu* DNA Ligase (4-0.001U) per reaction. Tet curves (marked in green) correspond to the amplified ligation product. Four out of five Tet curves are tightly grouped. This means that the amount of ligation product was the same. Only the curve corresponding to 0.001U of ligase per reaction rises one cycle later.

matched DOs are ligated. At longer ligation times the match/mismatch ratio begins to decrease. To provide a quantitative ligation of slow-hybridizing DO sets, it is necessary to increase the ligation time; however, to prevent mismatch ligation of fast-hybridizing DOs, it is desirable to decrease it. Sub-cycles are a compromise settlement: 3 minutes, the total ligation time, which is sufficient for slow-hybridizing DO sets, is divided into six 30 second intervals, which are short enough to maintain an appropriately low background for fast-hybridizing DO sets.

PEG 6000 is an essential component of the ligation mixture. In the presence of 15% PEG 6000, ten-fold lower concentrations of DOs may be used in ligation reactions. Figure 23 shows that in 2 hours, 1 amol of artificial template (this corresponds to ~50ng of *A. thaliana* genomic DNA) is completely hybridized with DOs at concentrations ≥63.2pM without PEG 6000 and at ≥6.3pM with 15% PEG 6000.

PEG 6000 concentration above 5% inhibits the PCR amplification (data not shown). The optimal concentration of PEG 6000 in a ligation reaction lies in the range of 7.5-20%. We use 15% PEG 6000, which is diluted to 3.75% after the addition of amplification mixture.

Concentration of the detector oligonucleotides is one of the most important parameters of the LDR. An insufficient amount of DOs leads to suppression of the signal due to incomplete hybridization to the target during the ligation cycle; an oversupply of DOs results in an increase in background due to (i) false ligation and (ii) false amplification of unligated DOs that have regions corresponding to the PCR primers.

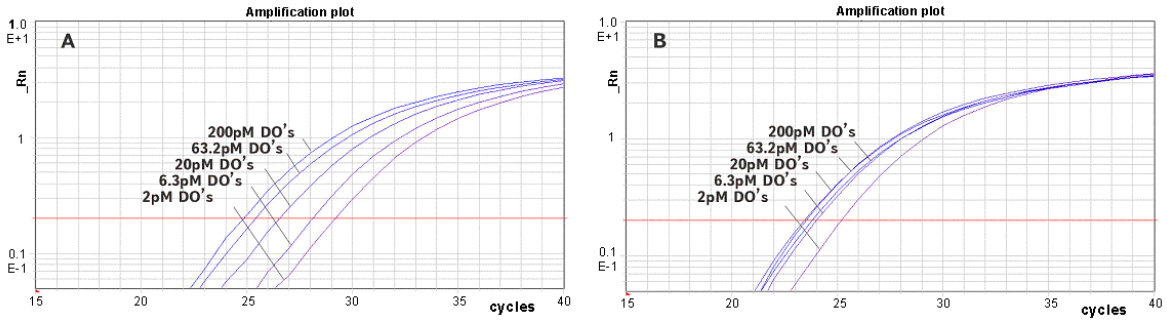

Figure 23. Real-time PCR curves for amplification of ligation products obtained after 2 hours of hybridization of DOs at different concentrations (2, 6.3, 20, 63.2 and 200pM) to 1 amol of template without PEG 6000 (A) and in the presence of 15% PEG 6000 (B).

Figure 24 shows that 10pM - 1nM concentrations of DOs provide quantitative hybridization with the target in the cyclic ligation (20 seconds for ligation per cycle) in the presence of 15% PEG 6000. Although the fluorescence signal is about two-fold less at 10pM compared to 1nM, it is sufficient for reliable detection.

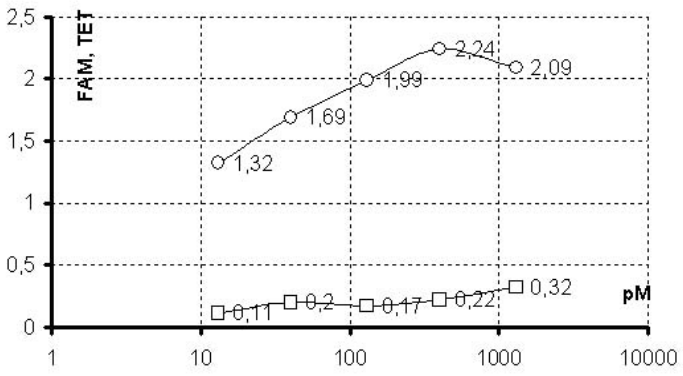

Figure 24. Influence of DO concentration (13pM, 40pM, 130pM, 400pM, and 1300pM) on the fluorescence signal in LDR. Genotyping was performed on 50ng of L*er* DNA. Tet and Fam signals are shown as circles and squares respectively.

Different DOs may have different minimal concentrations because of unique secondary structures and sequence-specific interactions with non-target genome sequences. We have selected the following concentrations of DOs for routine genotyping: 100pM for *A. thaliana* genomic DNA and 250pM for human DNA.

LDR-TaqMan genotyping kits

On the basis of the LDR-TaqMan technology, two ready-to-use genotyping kits were prepared: the *Arabidopsis thaliana* whole genome kit for 138 SNP loci polymorphic between accessions Columbia (Col-0) and C24, and the human genotyping kit for 9 clinically important SNPs in loci *LCT, APOE, SDF1, CCR5, RANTES, FAAH, AGT*, and *MTHFR*.

Preparation of enzymes and Detector oligonucleotides

With commercial reagents and consumables, the price of a single genotyping reaction in the "one tube – one locus" procedure is about 0.8€ (Table 5). For multiplex procedure the expenses are lower as a result of performing ligation in one tube for several loci. For example, a 30-plex ligation reduces the genotyping price to ~0.65€ per locus.

Table 5. Estimation of the price for a single genotyping by the "one tube-one locus" procedure (work expenses not included) using commercially available or homemade components.

Components	Price per genotyping reaction	
	Commercial components	Homemade components
Pfu ligase	0.1€ (Stratagene)	~0.001€ (home-made)
96-well reaction plate	0.023€ (Corning, ABgene)	0.023€ (Corning, ABgene)
TaqMan reagents	0.505€ (Applied Biosystems)	~0.002€ (home-made)
TaqMan probes	~0.075€ (Applied Biosystems)	~0.075€
96-well optical reaction plate and optical cover	0.090€ (Applied Biosystems)	0.023€ (Corning, ABgene)
PCR primers	~0.001€	~0.001€
	Total: ~0.8€	Total: ~0.15€

The main costs of the method are the enzymes (*Pfu* ligase and hot start *Taq* polymerase), the TaqMan probes, and the optical plates. Homemade enzymes and cheaper plastic, which proved to be adequate for fluorescence analysis, allowed us to

reduce the price to ~0.15€ and ~0.10€ per genotyping reaction in the "one tube – one locus" and the "one tube – many loci" procedures respectively.

Assembled ready-to-use genotyping kits included:

- *Pfu* ligase and 10x ligase buffer;
- 1.25x TaqMan amplification mixture including TaqMan probes, PCR primers, and Hot Start *Taq* polymerase;
- DO sets for all loci;
- Locus-specific PCR primers.

Homemade Pfu DNA ligase

Pfu DNA ligase is the key enzyme of the LDR-TaqMan method. It was selected from five commercially available thermostable ligases because it exhibited the highest ligation specificity.

We have cloned the ORF of the *Pfu* DNA ligase and prepared two expression constructns (for wild type and for His_6-*Pfu* ligase). Both wild type and His_6-*Pfu* proteins were purified according to the elaborated procedure (described in "Materials and methods"). Typically, ~$2x10^6$U of enzyme was obtained from 1L of the expression culture.

Both *Pfu* and His_6-*Pfu* ligases were indistinguishable from the commercially available enzyme (Stratagene) in terms of specific activity (~200U/μg), thermostability (half life time at 96°C is ~90 minutes), and specificity.

Homemade hot start Taq DNA polymerase

A hot start is required for the TaqMan assay in the LDR-TaqMan procedure. Normal start or manual hot start showed an inacceptibly high background. Therefore, we have prepared hot start *Taq* polymerase by chemical inactivation of the enzyme with formaldehyde [Ivanov et al., 2001]

The *Taq* polymerase expression construct was prepared as previously described [Engelke et al., 1990]. The purification protocol is a modified version of that suggested in the article. Usually, from1Lr of culture, ~1-$2x10^6$ U of enzyme were obtained.

Formaldehyde treatment was performed according to a modified protocol described previously [Ivanov et al., 2001]. The modifications allowed us to make the procedure faster and easier (see “Materials and methods”). The typical yield of hot start *Taq* polymerase is 1-5% from the original activity of the enzyme (from 1000U of *Taq*

polymerase, we get 25U of hot start enzyme). The activation level (activity after/before heating at 95°C for 15 minutes) is estimated to be 5-10x10^3 times.
In the LDR-TaqMan reaction, homemade hot start *Taq* polymerase is indistinguishable from the commercial enzyme (AmpliTaq Gold from Applied Biosystems).

Plastic consumables

Special optical plates and optical covers are used for fluorescence detection. However, common 96- and 384-well reaction plates (not marked "optical" and about one-fourth the price) also provide reliable signal detection. Optical covers might also be substituted for cheaper alternatives; for example, some types of common office tape and some non-optical covers for PCR plates were found to be appropriate for fluorescence analysis. It is therefore expedient for the end user to check different available plates and sealing tapes for genotyping. Plates and sealing material may be initially tested using a hand-held UV lamp: they should not fluoresce.

Plant and human genomic DNA isolation using homemade spin columns

We have previously published a protocol for plasmid DNA purification using homemade silica spin columns [Borodina et al., 2003a]. Since homemade columns have proven to be at least as convenient as commercial analogues, and also much cheaper, we have adjusted the protocol for isolation of genomic DNA from plant material and saliva. The average DNA yield was ~3µg from 100mg of leaves and ~50µg from 2ml of saliva collected using the "swish and spit" method [Hayney et al., 1995]. This is comparable to results obtained with the DNeasy Plant Kit (Qiagen), and with the phenol/chlorophorm isolation method for human DNA.

Genomic DNA isolated using homemade spin columns is suitable for SNP genotyping by the LDR-TaqMan method.

Ligation-based synthesis of Detector Oligonucleotides (DOs)

In addition to the genotyping price, there is also a setup price for each locus that must be considered when estimating costs. Three detector oligonucleotides and one PCR primer (for the multi-locus procedure only) are required for each locus.

DOs are rather long: 65-80nt for "left" DOs and 50-60nt for "right" DOs. The distinct block structure of DOs i.e. the combination of parts corresponding to constant

regions (Region I and II for DORs, and Region V for DOLs) and variable regions (Regions III and IV) allows for their cost-effective generation by ligation-based synthesis (LBS) [Borodina et al., 2003; Soldatov et al., 2004].

We have developed the LBS for synthesis of padlock probes. The LDR-TaqMan method was tested first with padlocks that were then changed to linear DOs. Padlocks are 90-150nt oligonucleotides that consist of locus-specific regions on both 3'- and 5'-ends, which are connected by a universal linker part. They are expensive and difficult to prepare. The yield and quality of the conventional phosphoramidite synthesis decrease dramatically when the length of an oligonucleotide is more than 60nt. For 80-100nt long padlocks, a double purification (HPLC and PAGE) is required. Oligonucleotide-producing companies cannot guarantee yield for oligos longer than 100nt and do not take orders for primers longer than 130nt.

Figure 25A illustrates the principle of assembling a padlock by LBS: three oligonucleotides (#a, #d and #b in Figure 25) are ligated to each other by T4 ligase, using two adaptor oligos (#c1 and #c2). Full-length padlocks are purified through denaturing PAGE (Figure 25B).

For synthesis of padlocks, LBS has by now no better alternatives considering quality and yield. Two PCR-based methods for padlock synthesis were suggested recently [Antson et al., 2000; Myer and Day, 2001], but both are laborious and expensive compared to the LBS technique.

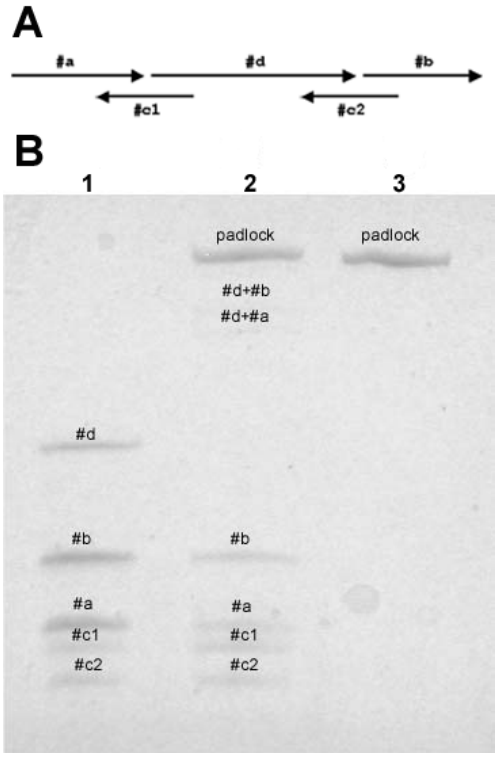

Figure 25. Preparation of padlock probes. **A.** Scheme of ligation. **B.** Step by step PAGE analysis of padlock synthesis. Three identical preparations of padlock probes were loaded in the gel at different stages: (1) before ligation, (2) after ligation and (3) after ligation and PAGE purification.

For preparation of DOs, LBS also provides noticeable benefits in terms of price and quality of oligonucleotids. A 5 nmol-scale LBS preparation of a DO set for one locus (three DOs each ~70nt long) costs ~70€. The same amount could be obtained from an oligonucleotide-synthesis company for ~130€. The more DOs are prepared, the lower the price will be per DO, since common parts have to be synthesized only once. Using homemade T4 DNA ligase (see "Materials and methods") makes the procedure even cheaper.

The alignment of the oligonucleotides involved in LBS of a DO set for the locus "01" is shown in Figure 15C. Pairs of oligonucleotides (#L1 and #L1_01, #L2 and #L2_01, #R1 and #R_01) are annealed to adaptor oligonucleotides (#aL1, #aL2 and #aR, respectively) and ligated. 5nt overlaps of locus-specific oligonucleotides with adaptor oligonucleotides were found to be sufficient for efficient annealing and ligation. Oligonucleotides #L1, #L2, #R, and adaptor oligonucleotides #aL1, #aL2 and #aR are common for all DO-sets. They were ordered only once and do not have to be repeatedly synthesized for every DO.

In 1 nmol-scale LBS, approximately 700 pmol of DOs are obtained, which is enough for ~10^6 LDR-TaqMan genotyping reactions. An example of preparation of the DO set for *A.thaliana* SNP locus “2” is shown in Figure 26.

The 155 sets of DOs (three DOs per locus) used in this work were prepared by LBS. The method proved to be robust and convenient and requires no locus-specific optimization.

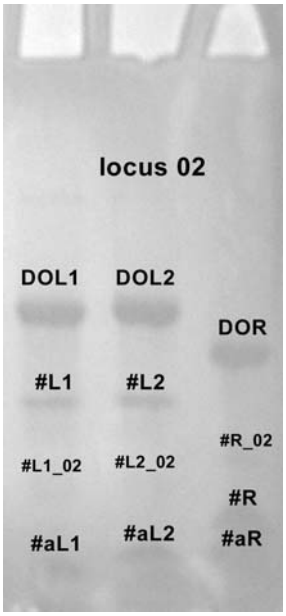

Figure 26. Preparation of the DO set for *A. thaliana* SNP locus "2" by LBS. After ligation, full-length DOs are separated from other oligonucleotides in 10% denaturing PAGE. Essentially all locus-specific oligonucleotides are consumed. Adaptor oligonucleotides #aL1, #aL2 and #aR remain intact; traces of other oligos involved in the reaction may be seen.

Arabidopsis thaliana SNP kit

Arabidopsis thaliana is an important model organism for plant research [Meinke et al., 1998]. The available genomic sequence of *A. thaliana* [The Arabidopsis Genome Initiative (AGI), 2000] provides a platform for functional studies and also for revealing the genetic background of complex phenotypic traits.

Most genetic studies in *A. thaliana* involve mutants, crosses, and mapping populations derived from the genetically distinct Columbia (Col-0) and Landsberg *erecta* (L*er*) accessions. Genetic variation harbored by other accessions of *A. thaliana,* which is also an important source for identifying functionally important genes [Alonso-Blanco and Koornneef , 2000; Steinmetz et al., 2000], remains unexplored.

The "Genomanalyse in Biologischen System Pflanze (GABI) - Max-Planck *Arabidopsis* Consortium" (MASC) was created in order to characterize genomic variations in different accessions of *A. thaliana* and to create an efficient genotyping tool for this flowering plant.

Within the framework of this project, our collaborators have identified ~8000 SNPs in 12 *A. thaliana* accessions [Schmid et al., 2003] and assembled a 112-marker set for genotyping of Col-0/C24 hybrid lines using the commercially available SNaPshot assay [Törjek et al., 2003]. Our group has developed the LDR-TaqMan SNP-genotyping method and prepared a whole-genome kit for 138 SNP loci (Col/C24), located between the SNPs selected for the SNaPshot assay. The kit has been transferred to the GABI-MASC community for routine genotyping.

SNP loci for the LDR-TaqMan SNP genotyping kit were selected from the GABI-MASC SNP database [Schmid et al., 2003; www.mpiz-koeln.mpg.de/masc/]. Guiding criteria for the selection of SNP loci were as follows:

- desired spacing ~1Mbp;
- SNP location between polymorphisms selected for SNaPshot-based genotyping;
- uniqueness of the SNP-containing sequence (~30nt from both sides of the SNP position); there are multiple duplications in the *A. thaliana* genome [Arabidopsis Genome Initiative (AGI), 2000], which might cause invalid results;

Tandem repeats and long single base stretches in the SNP-adjacent sequence were avoided when possible.

138 selected loci are listed in Figure 27b. They are evenly distributed throughout the genome with an average physical distance of 0.86Mbp between them. Genomic sequences (~30nt from both sides of the SNP position) are given in Supplement 1 (for loci 1-138). The approximate distribution of these SNPs with respect to those previously used by our collaborators is shown in Figure 27a.

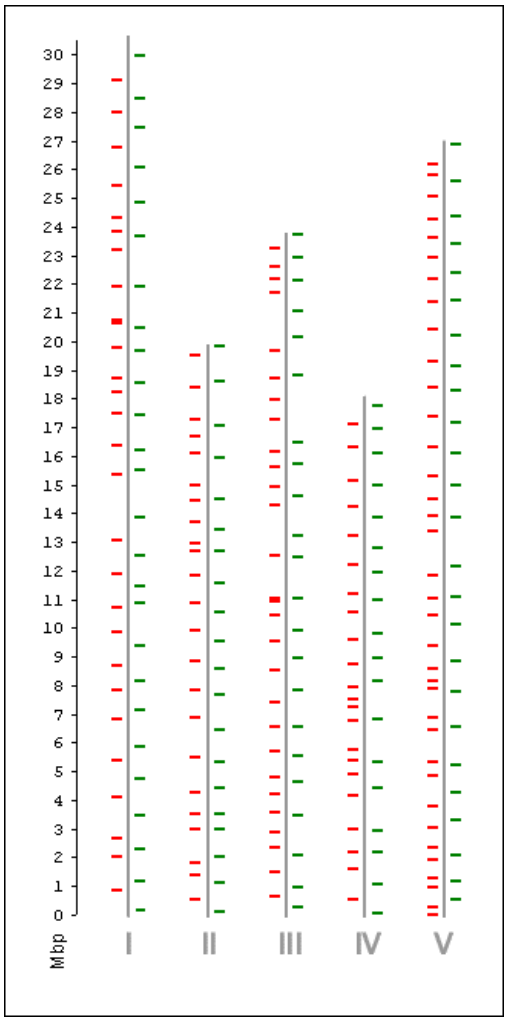

Figure 27a. Relative distribution of the 138 SNP loci used in the LDR-TaqMan kit (red dash lines) and the 112 SNP loci chosen for SNaPshot analysis (green dash lines) along the five chromosomes of *A. thaliana*.

DOs for all selected loci were prepared by LBS. Locus-specific oligonucleotides (Supplement 2) were selected according to the algorithm described in "Materials and methods". Allele-specific sequences corresponding to #fam and #tet were assigned to Col and C24 SNP variants, respectively. All DO sets were monitored for correct results in genotypings on Col, C24 and heterozygous DNA samples. No locus required specific optimization. Since some SNPs in the database were not verified, the polymorphic state of 10 SNPs (MASC01909, MASC05842, MASC05550, MASC00564, MASC04881, MASC04672, MASC06872, MASC03231, MASC01615, MASC01060) could not be confirmed; they were substituted by closely located loci.

To test the kit, blind genotyping of all 138 loci was performed for 10 DNA samples isolated from Col/C24 hybrids. On average, 5ng was taken per genotyping. From 1380 genotypings, 16 had to be repeated because of technical problems (drying of the reaction, or detection errors).

The LDR-TaqMan genotyping results for 138 SNP loci were compared to those obtained with the SNaPshot assay (112 SNP loci). These techniques displayed an identical distribution of Col-0, C24 and heterozygous genomic regions of the plants analyzed. Genotyping results for SNP loci located on the 1st chromosome are presented in Figure 28.

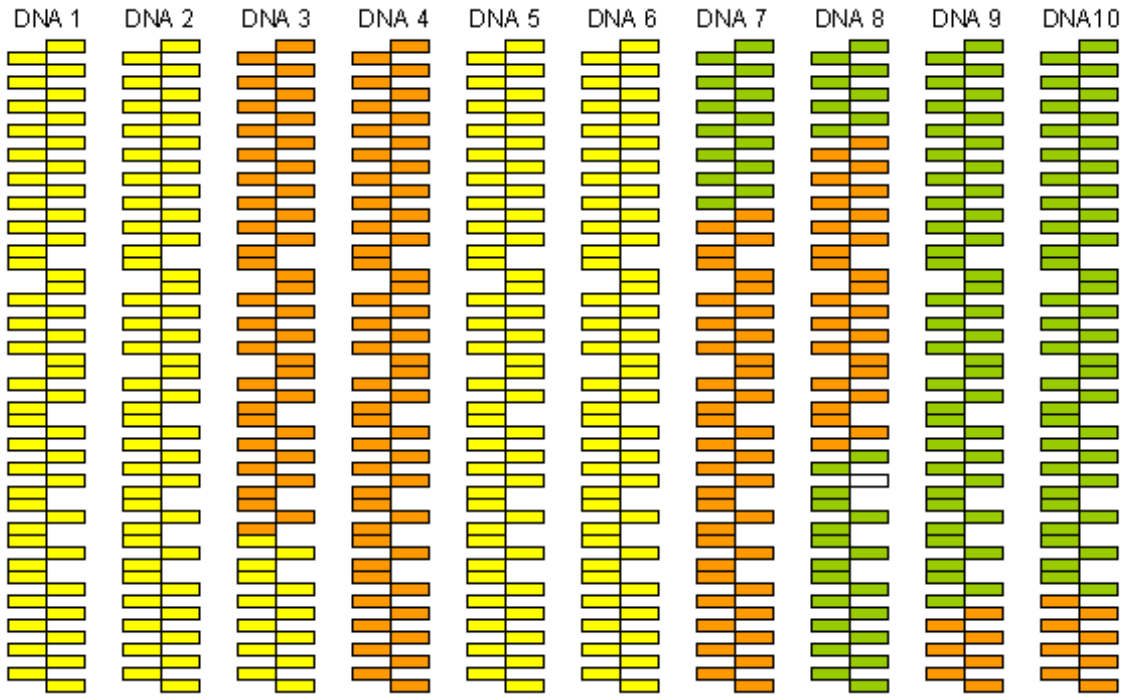

Figure 28. Genotyping results for the SNP loci located on the I[st] chromosome. Genotyping was performed by LDR-TaqMan (left side of each column) and SNaPshot (right side) methods on 10 DNA samples (DNA 1-10). SNP loci are represented by rectangles, arranged according to relative positions. Col-0 variants are marked in yellow, C24 in green, and heterozygous loci in orange. Empty rectangles correspond to loci, for which no SNaPshot data is available.

Genotyping results for three SNP loci located on the borders of the genomic regions contradicted the results obtained by our collaborators. LDR-TaqMan genotyping of the problem loci was repeated twice and gave the same results.

Signals obtained were univocal for all genotypings. Figure 29 illustrates the distribution of end point signals, grouped for SNPs located on one chromosome. Even

taken together for all DNA samples, signals for homo- and heterozygotes are clearly grouped and well separated from each other.

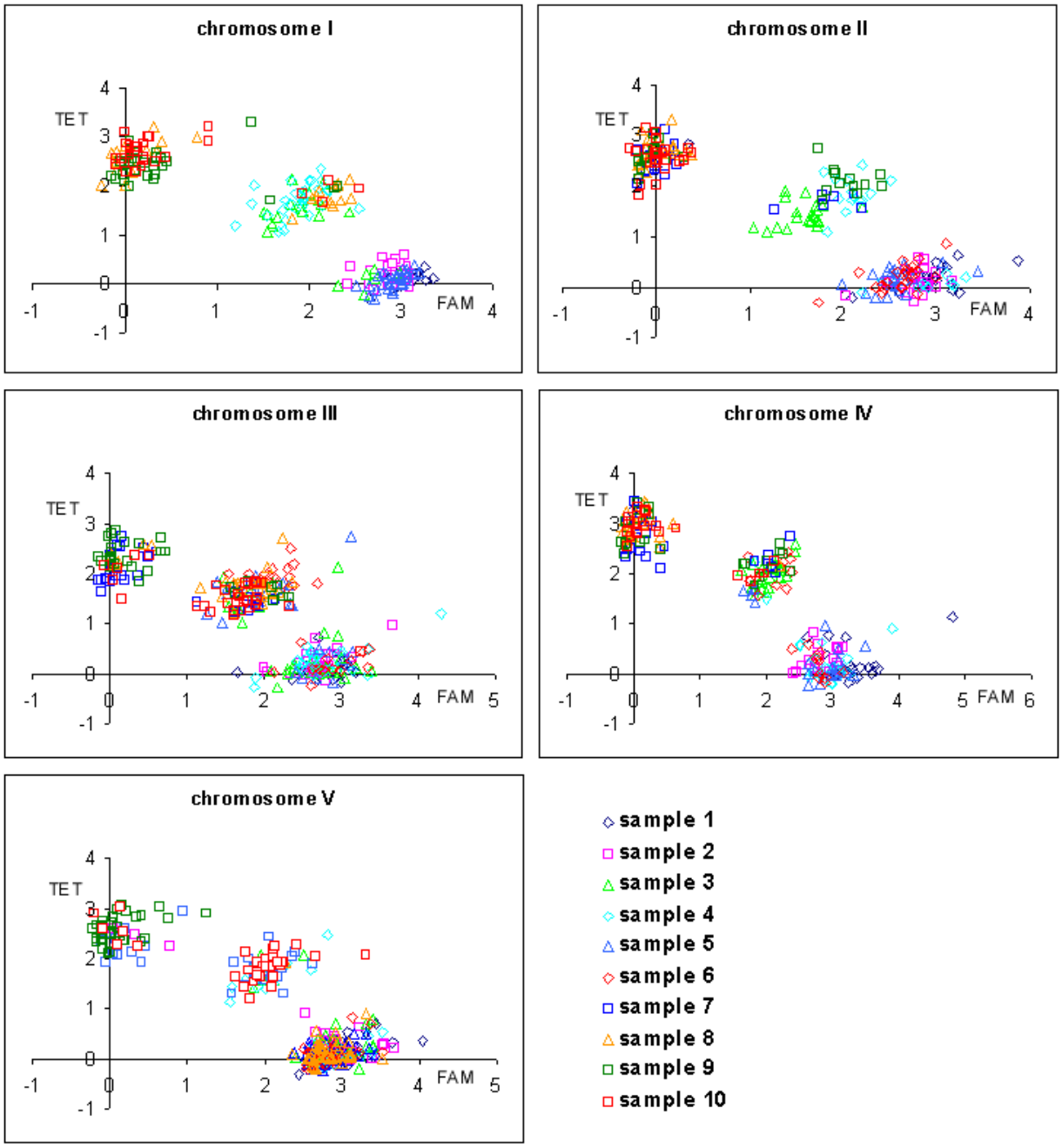

Figure 29. End point signal distribution for 1380 independent genotypings on 10 *A. thaliana* genomic DNA samples.

Human SNP kit

The objective of the joint German-Russian program was to reveal spectrum and frequencies of alleles influencing health in human populations of Eurasia.

Nine SNP loci listed in Table 6 were the focus of this study.

Table 6. Human polymorphic loci analyzed in the study.

Number	Locus	Polymorphism major/minor allele	Clinical value
1	*LCT*	C/T	marker for diagnosis of adult hypolactasia
2	*CCR5*	+/delta32	associated with resistance to HIV-1 infection
3	*SDF1*	G/A	associated with resistance to HIV-1 infection
4	*RANTES*	G/A	associated with an increased frequency of HIV-1 infection
5	*ApoE*	C/T	associated with development of atherosclerosis
6	*FAAH*	C/A	associated with drug/alcohol use
7	*AGT*	C/T	associated with blood pressure
8	*AGT*	T/C	associated with blood pressure
9	*MTHFR*	C/T	associated with cancer

(1) ***LCT (–13910 C/T)***

A C to T change residing 13910bp upstream of the lactase (*LCT*) gene at chromosome 2q21-22 shows complete association with lactase persistence [Enattah et al., 2002]. The genotype C/C_{-13910} in a homozygous form is associated with lactose malabsorption (LM) and genotypes C/T_{-13910} and T/T_{-13910} with lactose absorption (LA). This SNP may be used as a clinical marker for diagnosis of adult hypolactasia.

(2,3) ***CCR5 (+/delta32), SDF1 (-801G/A)***

CCR5 (beta-chemokine receptor 5) and *SDF1* (stromal-derived factor) genes play critical role in resistance to human immunodeficiency virus type 1 (HIV-1) infection. A homozygous 32nt deletion in the *CCR5* gene leads to complete resistance to HIV-1 [Dean et al., 1996; Huang et al., 1996] and heterozygous individuals display delayed onset of the acquired immunodeficiency syndrome (AIDS) [Liu et al., 1996].

The common polymorphism *SDF1*-3'*A (G/A)* is located in an evolutionarily conserved segment of the 3'-untranslated region of the *SDF1* gene. In the homozygous state, *SDF1*-3'*A*/3'*A* also delays the onset of AIDS [Winkler et al., 1998]

(4) ***RANTES (-403 G/A)***

The -403 G/A SNP in the promoter of the *RANTES* (regulated on activation normal T cell expressed and secreted) gene is associated with an increased frequency of HIV-1 infection. An A in the SNP position results in diminished transcription of *RANTES*, which leads to accelerate progression of AIDS [An et al., 2002].

(5) ***ApoE (3932 C/T)***
Apolipoprotein E (*ApoE*) is a ligand for lipid metabolism–related receptors that bind several lipoproteins. *ApoE (3932 C/T)* is one of the SNPs associated with plasma lipid metabolism and development of atherosclerosis [Ghiselli et al., 1982; Davignon et al., 1988].

(6) ***FAAH (385C/A)***
A C to A single nucleotide polymorphism at position 385 in the human *FAAH* (fatty acid amide hydrolase) converts a conserved proline residue in *FAAH* to threonine (P129T). This SNP is strongly associated with both street drug use and problem drug/alcohol use [Sipe et al., 2002], suggesting a potential role for the FAAH-endocannabinoid system in regulating addictive behavior [Chiang et al., 2004].

(7,8) ***AGT (M174T, M235T)***
AGT (angiotensinogen) gene polymorphisms T174M (C/T) and M235T (T/C) show significant associations with average systolic (SBP) and diastolic blood pressure (DBP) [Robinson and Williams, 2004; Rotimi et al., 1997], and the extent of coronary heart disease [Gardemann et al., 1999].

(9) ***MTHFR (A222V)***
MTHFR (5-methylenetetrahydrofolate reductase) participates in the metabolism of folate. A C/T polymorphism at 677 position of the *MTHFR* gene correlates with reduced enzyme activity and increased thermolability [Frosst et al., 1995]. A link has been found between this polymorphism and the tumor response to fluoropyrimidine-based chemotherapy; thus, *MTHFR* genotyping may be of predictive benefit in selecting treatment regimens [Cohen et al., 2003].

Our partners from the Institute of General Genetics (Russian Academy of Sciences, Moscow) were to collect DNA samples from ~100 individuals from each of 20 Eurasian populations of European and Asiatic origin (Russian, Ukrainian, Belorussian, Moldavian, Gagause, Bashkir, Tatar, Chuvash, Udmurt, Mari, Uigur, Uzbek, Kazakh, Lackh, Avar, Lezgin, Cherkess, Nogai, Kumyk, Kalmyk) and to characterize them.

Our task was to provide them with a ready-to-use SNP genotyping kit.

Genomic sequences adjacent to the polymorphic loci (Supplement 5) were derived from the NCBI database. DO sets for all nine loci were prepared by LBS. Locus-specific oligonucleotides were designed according to the algorithm in "Materials and methods" (Supplement 6). Allele-specific sequences corresponding to #fam and #tet were assigned respectively to the major and minor allelic variants (mentioned as major/minor in the gene list). All DO sets were monitored for accuracy in genotypings on homozygous and heterozygous DNA (genotypes were verified by sequencing). None of the loci required specific optimization.

The kit was used for determining allelic frequencies of all 9 loci in Ukrainians (115 samples) and Belorussians (101 samples). *LCT (–13910 C/T*) was also genotyped in Russians (97 samples).

Genotyping data obtained with the kit was used in the analysis of the distribution of the HIV-protective alleles *CCR5delta32* and *SDF1*-3'*A* in Belorussians and Ukrainians (Tables 7 and 8) [Kozhekbaeva et al., 2004].

Table 7. Genotype and allele frequencies of the *CCR5* locus in the populations analyzed.

Population	Number of samples	Genotype frequency, %			Allele frequency (± standard error)	
		+/+	+/del	del/del	+	del
Ukranians	117	76.9	22.2	0.9	0.88±0.02	0.12±0.02
Belorussians	125	77.6	20.0	2.4	0.88±0.02	0.12±0.02

Table 8. Genotype and allele frequencies of the *SDF1* locus in the populations analyzed.

Population	Number of samples	Genotype frequency, %			Allele frequency (± standard error)	
		G/G	G/A	A/A	G	A
Ukranians	103	67.0	30.1	2.9	0.82±0.03	0.18±0.03
Belorussians	127	52.0	43.4	4.7	0.74±0.03	0.26±0.03

Genotyping was performed on 5-50 ng of human DNA, because the available amount of genomic DNA varied significantly between samples. The 10-fold difference had practically no effect on the end point fluorescence.

Among 1035 genotypings, 16 could not be performed with immediate success. 10 failed due to technical problems (pipetting mistakes or plate/cover defects) and were repeated, and 6 showed insufficient discrimination, presumably due to problems with the DNA quality. Genotyping results for *CCR5* and *SDF1* loci in Ukrainians were confirmed by conventional sequencing. The results of both approaches showed 100% coincidence. *LCT* genotyping results for Russians were compared to those obtained by restriction fragments length polymorphism (RFLP) technique, the method that was used for genotyping by our Russian partners). In 15 out of 97 samples, the results of the two

methods differed. Problem samples were also genotyped by sequencing, and in all cases, the results of the LDR-TaqMan method were verified.

An example of signal distribution in (Figure 30) illustrates the genotyping results for 9 loci in 115 Ukrainian DNA samples.

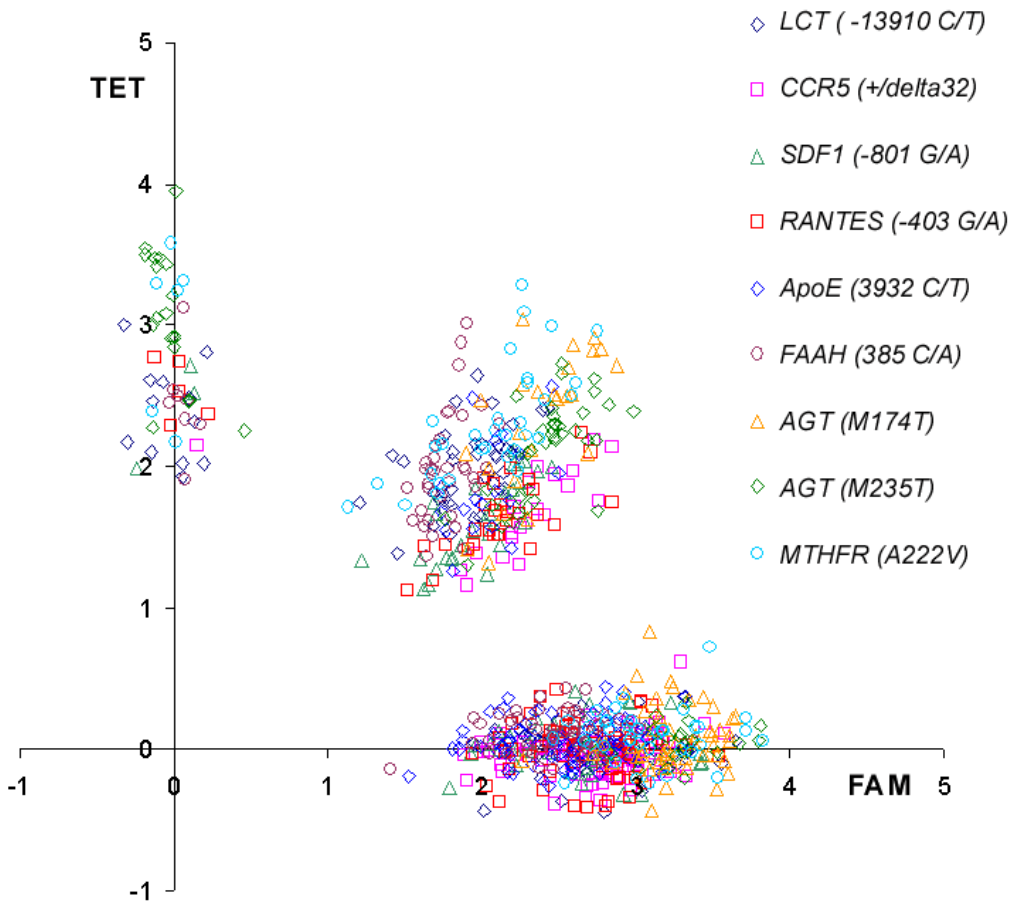

Figure 30. End point signal distribution for 1035 independent genotypings on human genomic DNA (Ukrainian samples).