

Optimized amplification of multicopy Y-chromosome sequences for the improved quantitative measurement of fetal male DNA in plasma by real-time PCR

Bernhard Zimmermann, Wolfgang Holzgreve, Sinuhe Hahn
Laboratory for Prenatal Medicine, University Women's Hospital, Basel, Switzerland.

Abstract:

The quantitative measurement of fetal DNA levels in maternal plasma has opened a new approach for the noninvasive prenatal diagnosis. One of the main applications is the prediction of fetal sex by the real-time quantitative PCR technique, as in the assessment of X-linked disorders. The amplification of Y-chromosome specific sequences is also the main research tool applied for the determination of fetal DNA levels in correlation with pregnancy related disorders, including common trisomies, preterm labor and preeclampsia. Even though it is only applicable to 50% of the samples since it is restricted to male pregnancies, this procedure is very straightforward and simple to perform. The precision and reproducibility of the quantitative measurements can greatly be improved by usage of Y-specific sequences that are present in more than one copy per Y-chromosome. The possibility to use an increased number of amplification targets per genome equivalent (GE) has been identified for a long time, but it has not been widely exploited.

Introduction

Since the discovery of circulating fetal DNA in the maternal circulation by Thomas in 1995, the real-time PCR amplification of DNA extracted from maternal plasma has emerged as the main strategy for the development of non-invasive methods for prenatal diagnosis. It is successfully used for the prenatal diagnosis of fetal sex and rhesus D status (genetic loci that are not present in the maternal genome) which are already being offered clinically. In addition to the relatively facile detection of these loci, the quantification of the fetal DNA in maternal plasma is currently being investigated as screening tool for pregnancy related pathologies. But the accurate quantification of genetic material from maternal plasma is not as straightforward as its mere detection. As we could observe in a recent comparison study between five laboratories, even use of the identical sample material and implementation of the same DNA extraction and real-time quantitative PCR procedures can result in major differences of the finalized data. In the comparison of numerical values from different studies using different protocols, major discrepancies can be expected. In the context of clinical application it will be important to standardize several procedural steps which can lead to variability of the quantitative data. Also, a more precise description of the quantification procedures needs to be included when reporting results. Here we present an approach to address one of the major hindrances in the generation of precise and relevant data, the scarce nature of the material to be quantified. We try to highlight several important points to generate precise and accurate data.

We amended an existing real-time quantitative PCR protocol specific to the multicopy locus *DYS14* and evaluated it's performance in comparison to the currently used *SRY* targeting protocol. Our evaluation shows that the new assay performs with an increased precision and sensitivity. Furthermore we present possible approaches for the standardized scoring of low template amplification to exclude false positive results.

Methods

Primers and probes from the existing real-time PCR protocol for *DYS14* were redesigned to amplify sequences present in multiple copies on the Y-chromosome with 100% match. The real-time qPCR was performed in the ABI PRISM 7000 Sequence Detection System (Applied Biosystems, ABI, Switzerland). DNA in volumes of 1 to 4 µl was amplified singleplex in 25 µl reactions containing 300 nM of each primer (HPLC purified, Microsynth, Switzerland), 200 nM of the MGB probe (ABI), and at a 1 x concentration of the TaqMan Universal PCR master mix (ABI). For subsequent melting curve analysis, 2 µl of the SYBR Green TaqMan Universal PCR master mix (ABI) were added to the reaction and a melting curve was performed on the SDS7000.

In reactions without probe, the SYBR Green TaqMan Universal PCR master mix (ABI) was used.

The reactions were prepared on ice and cycling conditions were as follows: 2 min at 50°C and 10 min at 95°C, followed by 40-50 cycles of 1 min at 60°C and 15 s at 95°C.

Quantities are presented in genome equivalents (GE), the amount of DNA present in one cell.

target	name of primer	sequence	length	100% BLAST hits
DYS14	DYS14_F1713	CATCCAGAGCGTCCCTGG	18	10
	DYS14_F_BZ	CCAGAGCGTCCCTGGCT	17	10
	DYS14_F_BZ17	GGGCCAATGTGTATCCTCTCT	22	9
	DYS14_F_BZ23	GCTTCTGGGCAATGTGTATC	22	11
	DYS14_R880	TTCCCTTTTGTCCCAAA	19	0
DYS14_R_BZ	TCGCCCTTTGTCCCAAG	19	8	
DYS14_R_BZ4	TCACCTACACTTCCCGAGAGC	22	8	
DYS14_R_BZ5	GCCCATGGCTACTACACTTC	22	8	
DYS14_R_BZ22	ACGCCGAGCTGCCAT	16	10	
SRY	SRY_F	TCGCTAAAGAAACCGTGGAT	21	1
	SRY_R	AGATAATGTTGTTGCTAAGACTGGAT	26	1
name of probe				
DYS14_MGB	TCTAGTGGAGAGGTGCTC	18	9	
DYS14_TAMRA	CGAAGCCGAGCTGCCATCA	20	1	
SRY_MGB	TCCACCAACCTCTT	15	1	

Table 1: Analysis of sequences: To optimize the *DYS14* protocol we blasted the primer and probe sequences of the original assay (*DYS14_F1713*, *DYS14_R880* and *DYS14_TAMRA*) and the corresponding amplicon on the locus x06325 to the human genome. We identified one mismatch at base one of the reverse primer and at base 4 of the probe in alignment with most copies of the chromosome Y sequence, the original probe sequence has only one 100% match on the Y-chromosome.

Also primer dimer formation was checked by the Primer Express program. This analysis revealed very strong primer dimers for the original primers. Furthermore we identified a number of ambiguous bases in the target sequences, which were excluded from the sequences for the new primers.

Results

Testing of primer combinations

SYBR green amplifications with the original and new primer pairs produced false positive results for female DNA of high and low concentration (12000 and 12 GE per PCR). Melting curve analysis allowed the identification of the amplification products from the female DNA as unpecific. Discrimination was more evident for some of the new primer pairs. In parallel amplification, also the *SRY* primer pair, whose specificity in combination with a probe is well established, showed unpecific product formation. As expected, the *SRY* product has a much lower melting temperature due to its shorter length.

Reduction of primer concentration would be possible to avoid unpecific products, especially since primer dimer formation (as found in alignments) is the most probable cause.

However, since further analysis of selected primer pairs with the probe was specific, we decided to keep the primer concentrations high to guarantee optimal sensitivity. After melting curve analysis of the negative control reactions (female genomic DNA) that did not generate an amplification curve, unpecific product peaks could again be detected.

The primer combination *DYS14_F_BZ17* x *DYS14_R_BZ5* produced the lowest C_T values, optimal efficiency and melting temperatures of 76°C for specific and 73°C for unpecific product and was selected for the final assay.

TaqMan protocol Precision and Sensitivity

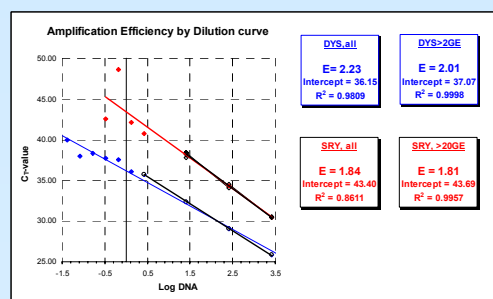


Fig. 1: Standard curves from the new assay in comparison to the *SRY*-specific assay: Each data point is the average of 4 replicate amplifications (2 for standards with >100 GE) of genomic DNA.

The average was calculated by converting C_T values into arbitrary quantities ($AQ_Ty = 10^{-(C_T/3.33)}$), averaging these and calculating an average C_T from these averaged quantities ($C_T = \log(AQ_Ty) * 3.33$). This procedure allows to include wells with no amplification in the standard curve.

For each assay two standard curves were generated: one including all dilutions, one with replicates of more than 200 target sequences per reaction.

The two standard curves are very similar. When very low concentrated samples are included in the standard curve, the correlation decreases and the use of too few replicates of low concentration could skew the data.

Relevant template sequence number for standard curves: We propose to use at least 10 template sequences per reaction for standard curves, as more dilute template concentrations decrease correlation coefficients and may lead to inaccurate quantification.

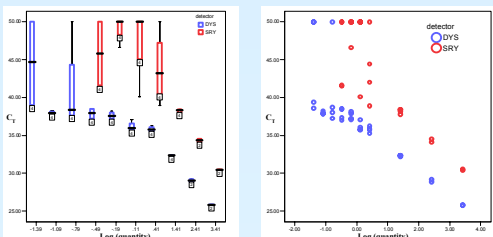


Fig. 2: Amplification of dilutions of genomic DNA: Results are presented as box-and-scatter-plot. Replicates with no amplification are assigned a C_T of 50, in order to represent the data in the graphs. Numbers below the box are replicate amplifications.

The *DYS14* assay is still able to generate precise data when the *SRY* amplification results in high C_T values of great variability and replicates without amplification.

The sensitivity of the *SRY* assay is confirmed to be close to 1 GE, the sensitivity of the *DYS* amplification is clearly superior with close to 0.1 GE.

Plasma samples

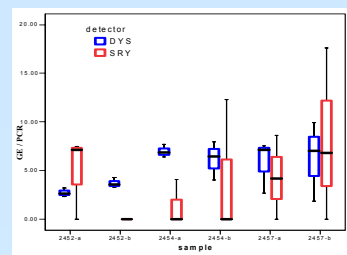


Fig. 3: Replicate extractions of plasma samples 2452, 2454 and 2457 quantified by real-time PCR for *DYS14* and *SRY*. 2µl DNA equivalent of 15 µl plasma. Replicate extractions separate (left) and combined (right).

For the triplicate reactions, the *DYS* data is much more precise (Coefficients of variation of 8-53%) than the *SRY* targeting amplification (71-141%), which also has one negative triplicate. The data from replicate extractions is equivalent for the *DYS* protocol, while the *SRY* quantification is not.

The precision and reproducibility at low template concentrations are clearly increased with the multicopy targeting protocol.

Conclusions

To avoid false positives (which we observed in multiplexed amplifications), we propose to use an arbitrary cut-off value, which determines when the amount of fetal male DNA is considered zero or "none detected". This can be:

- all replicates amplify (or 2 of 3)
- a maximal C_T value,
- a more quantitative cut-off: a target sequences number (between 0.5 and 2 targets per reaction for quantitative studies, 10 or more for clinical diagnosis). The fraction of a copy for a *SRY* specific PCR is a questionable quantitative result, since either one copy is present or not.

To avoid the scoring of false positives as observed in SYBR Green amplifications, a melting curve should always be performed.

The application of real-time quantitative PCR for the measurement of multicopy Y-specific sequences is with the reduced variability an optimal tool for the assessment of fetal DNA concentration in maternal plasma and for the investigation of elevated levels under a variety of pregnancy related pathologies.

- Published results are often at means of 20 GE/ml, for many reactions this is only around 1 template copy per PCR. The molecular sampling error can play a considerable role at these low template concentrations and is markedly reduced with the *DYS* protocol. The detection limit of the protocol performed should be stated clearly: For example 1 ml plasma is extracted into 50 µl, and the PCR is performed with 5 µl DNA. The detection limit is 1 copy per 100 µl plasma for the *SRY* protocol and 1 copy per 10-20 µl plasma for *DYS*.

• Standard curves should not include samples with less than 10 template sequence copies, since lower amounts result in variable copy numbers in the reaction. Samples of relevant concentrations (1-10 GE) can be included in *DYS* standard curves, while for the *SRY* deviations in C_T values might influence results.

• With the increased sensitivity, fetal sex determination may be possible at even earlier stages of gestation than reported with single copy locus specific assays. Treatment of adrenal hyperplasia could be totally prevented in at least some male pregnancies.

• Specificity remains to be proven in large scale analysis. Also, *DYS14* might pick up non-relevant DNA from previous pregnancies.